AD No. 36198
ASTIA FILE COPY

CARNEGIE INSTITUTE OF TECHNOLOGY
PITTSBURGH 13, PENNSYLVANIA

PROBLEMS IN SERVOMECHANISM SYNTHESIS

REPORT NO. 2

E. M. Williams
W. P. Caywood, Jr.
R. C. Lyman

Progress Report on Work Conducted Under
Office of Naval Research Contract
N7ori - 30306 and 30308

THIS REPORT HAS BEEN DELIMITED
AND CLEARED FOR PUBLIC RELEASE
UNDER DOD DIRECTIVE 5200.20 AND
NO RESTRICTIONS ARE IMPOSED UPON
ITS USE AND DISCLOSURE.

DISTRIBUTION STATEMENT A

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED.

Problems in Servomechanism Synthesis

Report No. 2

by: Richard C. Lyman
Project Engineer in Electrical Engineering

Progress

This is the second of a series of reports of an investigation of servomechanism synthesis under Navy Contracts N7ori30306 and N7ori30308. The investigation was initiated originally for the purpose of studying the application of magnetic amplifiers in servomechanism design. It rapidly became apparent, however, that (A) it has already been demonstrated that magnetic amplifiers are applicable to servomechanism problems in a wide variety of cases, (B) that such applications are commonly made on a rule-of-thumb basis and that, (C) a more refined application of magnetic amplifiers in servomechanism design cannot be separated from the general problems of servomechanism synthesis and that this project should start with the redefinition of the basic problem, that of an optimum criterion for synthesis on a realistic basis.

In the classical approach to servomechanism analysis it is assumed that (1) signals are of regular functional form, typically step functions, (2) such artificial criteria as phase margin or peak overshoot values are satisfactory as measures of synthesis success. More recently servomechanism approaches have been directed to consider (1) the statistical properties of signals, (2) realistic criteria both linear and nonlinear, and (3) the treatment of systems which in themselves are digital in nature, their utilization being effected by quantizing the analogue signal with a transducer when necessary. This is the direction undertaken in this particular study.

The results of this study to date are as follows:

a) A method of evaluating a linear servomechanism system in the light of any type of signals that can be statistically described, and any type of error

criteria for which a power series approximation can be written (with the common quadratic norm as one term of the series). This work was described as report n number one.

b) A method of prescribing the optimum transfer function for systems having certain specified time-varying characteristics, or for signals having certain specified time-varying properties (including random variations), as well as the best adjustment for first and second order systems for a set of commonly assumed, statistically described signals which also may have time variations. In some respects this work is an extension of Wiener's methods. This work is described herein as report number two of the series, and was submitted to Carnegie Institute of Technology by R. C. Lyman as a Dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering.

c) A method of deriving a transfer function for a magnetic amplifier starting with the B-H characteristics of the non-linear material. The method permits determining in addition to the linear terms the harmonic distortion terms of both the envelope and the carrier functions. This material is being readied for report in the near future.

d) A new mathematical-graphical method of solving problems in combinational switching circuits. The importance cannot be overemphasized of the use of magnetic amplifiers as components of digital circuits, in which their inherent non-linearities are utilized for pulse reshaping rather than being a severe liability. Such use can proceed only at a limited pace until the mathematics of the design of digital circuitry are better known and in convenient form. The new method is essentially ready for report.

1) Symposium on Statistical Methods in Communication Engineering, Berkeley Campus, University of California, Aug. 17-18, 1953, "Generalized Servomechanism Evaluation", Caywood, Lyman, and Kaufman.

2) I.R.E. 1954 National Convention, "Optimization of Servosystems", Lyman and Caywood, Session: Radio Telemetry and Remote Control III.

TABLE OF CONTENTS

	Page
SYNOPSIS	1
ACKNOWLEDGEMENTS	2
INTRODUCTION	3
The Problem	3
Choice of Criterion	4
The Statistical Approach	5
Time-varying Spectra	6
Time-varying System Parameters	6
The Scope of this Dissertation	7
PART I, DERIVATION OF A MEAN SQUARE ERROR SPECTRUM	9
PART II, THE OPTIMIZATION PROBLEM	14
The Method of Solution	14
Optimum, Idealized System Transfer Functions	18
Independently Time-varying Input Spectra	18
Dependently Time-varying Input Spectra	24
Time-varying System Closed Loop Gain	26
Optimization of Physically Realizable Systems	28
Optimization of Invariant Systems	31
Optimization of Random System Parameters	39
CONCLUSIONS	49
FUTURE INVESTIGATIONS	50
BIBLIOGRAPHY	51

TABLE OF CONTENTS

	Page
APPENDIX A, EXPERIMENTAL CHECK	52
Single Sine Wave Input	53
Analytical Solution	53
Experimental Check	54
Spectral Density Input	55
Analytical Solution	55
Experimental Check	56
APPENDIX B, OPTIMIZATION OF A FIRST ORDER SYSTEM	61
APPENDIX C, OPTIMIZATION OF A SECOND ORDER SYSTEM	66

List of Figures

		Page
Figure 1.	Input Amplitude Spectra	32
Figure 2.	Servosystems	34
Figure 3.	System Mean Square Errors	37
Figure 4.	Random System Parameter	43
Figure 5.	Comparison of Experimental Points and Analytical Curve for Single Sine Wave Input	58
Figure 6.	Experimental Setups	59
Figure 7.	Comparison of Experimental Points and Analytical Curve for Spectral Density Input	60
Figure 8.	Contour Map of the Mean Square Error Surface for System 3	65
Figure 9.	Contour Map of the Mean Square Error Surface for the Second Order System	69

SYNOPSIS

This dissertation derives optimum transfer functions for servosystems. For the case of time-varying signal and noise input spectra, the optimum system has zero phase shift and an attenuation given in terms of the input spectra and their variances. For the case of a time-varying system parameter, a similar conclusion results. For a physically realizable system of a specified order, a formulation of the general solution is given and is solved for six important cases.

ACKNOWLEDGEMENTS

The author wishes to express his gratitude to Dr. William P. Caywood, Jr., for his helpful suggestions which lead to the problem treated in this dissertation and his encouraging guidance in all stages of the investigation. The suggestions of the department readers, Dr. James B. Woodford and Dr. Louis Depian, regarding the final composition of this dissertation are greatly appreciated.

This work has been supported in part by Office of Naval Research contract number Navy #9 T. O. #6.

INTRODUCTION

The Problem

This dissertation considers the problem of the synthesis of optimum servosystems which are subjected to signal inputs contaminated with noise. Of primary interest are those problems which are characterized by time-variations in the system parameters and in the statistical properties of the signal and noise inputs. The criterion of system optimization used in this dissertation is the minimization of the mean square error in the response of the servosystem.

The method of solution presented herein is primarily carried out in the frequency-plane rather than in the time-plane. Hence the concept of the frequency components present in the system inputs, output, and error, as described by their spectral densities, is used rather than the concept of their functional form with respect to time. The method is limited to consideration of linear systems.

The literature contains a great amount of work (1b,2) which deals with the problem of the optimization of invariant systems subjected to inputs having invariant statistical properties. However, the optimization problem which is characterized by time-variations in the parameters of the system and in the statistical properties of the input has not been treated in the literature. This dissertation constitutes an original and useful contribution to the field of servosystem synthesis and analysis by formulating the

latter problem and presenting solutions of it.

Choice of Criterion

There are several reasons for the choice of minimum mean square error as a criterion of system optimization. The principal one is the possibility of transformation of the mean square error of a system response from the time-plane directly into the frequency-plane. That this is true has been shown many times in the literature (1). In addition, this property of direct equivalence is valid only for the mean square criterion, not for any other powers. Direct transformation is desirable in that system responses and criteria are most naturally interpreted in the time-plane, whereas mathematical analysis is generally handled more conveniently in the frequency-plane. That the mean square value of the error in the time-plane can be computed as the integral of error spectral density over all real, positive frequencies has been proven in the literature (1a). Another reason for this choice lies in its suitability for a class of problems including those which require minimization of the power losses resulting from the error in system response. For the case of a second order system subjected to step function signals, it is interesting to note the following result. Minimization of the average error power (mean square error) specifies a ratio of damping to critical of 0.5, which is a relative first overshoot of 0.164. Such a response is generally considered to be too oscillatory,

e.g. a region of 0.6 to 1.0 has recently been advised (2b).

The Statistical Approach

An important characteristic of the method of solution presented herein is the use of statistical properties rather than functions of time to describe the system inputs, output, and error. A system subjected to an input which is known as a function of time is merely a signal generator. For the input to contain any useful information for controlling the system, it cannot be known with certainty at all future instants of time. It may be described only by its statistical properties. These properties, which may be spectral densities or correlation functions, must be obtained by certain averaging operations performed on the inputs expected over some period of operation of the system being designed. Such statistical information can only lead to the statistical properties of the system error, such as the error spectral density which is used in this method. As a result of this statistical approach to the optimization problem, an average of the system performance is optimized rather than any particular characteristics of the system's response as a function of time. The optimum system may be considered as that system which achieves the best compromise between passing the signal spectrum (and necessarily some noise along with it) and attenuating the noise spectrum (and some signal along with it).

Time-varying Spectra

The term "time-varying signal spectrum", as used in this dissertation, is a description of a signal having two particular characteristics. First, its statistical properties remain time-invariant over periods of time of a sufficient length to properly establish those properties. They may be established by measurement of its spectral density or its autocorrelation function. Second, the statistical properties change either gradually or suddenly to various other values over the complete range of conditions for which the over-all system performance is to be optimized. As a result, successive measurements of those properties made over that range of conditions are found to differ one from the other. As a simplified example, the spectrum of a signal input to a servosystem may exhibit a peak at one frequency during the day and at a quite different frequency during the night. If the system is not desired to have a "day-night" operation selector in order to be optimum during each period, then its overall performance must be optimized. The set of signal spectra measured for day and for night is a sufficient description of the signal. Those spectra constitute the values of the time-varying signal spectrum. A time-varying noise spectrum must have the same characteristics.

Time-varying System Parameters

The time-varying system parameters may include amplifications, time constants (electrical or mechanical), and the like, which exhibit random changes in value. These random

changes may be tolerated in the system because their cause is unknown or because their correction is not feasible. They may result from internally or externally caused changes in the environment of the system. Some examples are changes in temperature, pressure, humidity, and power sources. Another possibility is fading signal reception. These random variations may also be interpreted to result from inaccuracies of manufacture of some system part where the problem is to optimize the total of the performances of many similar systems, each containing such a part.

The Scope of this Dissertation

There are two major parts in this dissertation. In Part I, the spectrum of the mean square error (the error spectral density) is derived in a manner which differs in several respects from that found in the literature (1c, 3). The form of the spectrum derived herein shows promise of being more useful in some experimental applications than the form presented in the literature, for a reason explained later in this paragraph. The signal and noise inputs to the system are herein described by the amplitude and phase components of their voltage spectra rather than by their power spectral densities as is commonly done. The only restriction is that the signal and noise enter the system at the same point; thus the system input is a signal contaminated with noise. Correlation between the signal and noise is herein described by the signal and noise amplitude spectra

and a relative phase angle which refers the absolute phase angle of the noise to that of the signal. This is in contrast to the use of cross-spectral densities in the literature (1c). The resulting form of the mean square error spectrum derived herein may be often more useful for experimental application than that presented in the literature for the following reason. The cross spectral densities are mathematical concepts involving products of the Fourier transforms of signal and noise and of their complex conjugates. Their experimental determination probably must be accomplished by transformation of experimentally measured cross-correlation functions. In contrast, the relative phase angle may generally be measured directly, as described briefly in Part I.

In Part II, the mean square error spectrum is used to solve the optimization problem, especially as characterized by random, time-varying behavior in the input spectra and system parameters. These random variations are treated by operating on the mean square error or on its spectrum with a probability theory operation which has been termed Expected Value in one reference (4a) and Mathematical Expectation in another (5a). This operation may be applied to either the form of the mean square error spectrum derived herein or the form commonly found in the literature. Use of this operation to allow for the effects of the random variations characterizing the problem, rather than neglecting such random behavior

permits the design of systemshaving less mean square error. Optimum but nonrealizable system functions are derived for several cases of random time-variations. A general solution for the optimum design of realizable systems of a specified order is formulated and solved for six important cases.

PART I

DERIVATION OF A MEAN SQUARE ERROR SPECTRUM

This part of the dissertation contains a derivation of the mean square error spectrum. For the purposes of this derivation, the signal and noise inputs to the system are represented by voltage spectral densities rather than by power spectral densities. A voltage spectral density may be considered to be composed of an amplitude spectrum and a phase spectrum, these indicating the relative potential of each frequency component present and the phase angle of each component. Both the amplitude and phase spectra may be determined by using the Fourier transform in cases where the signal may be described functionally in time. Otherwise, actual measurement using a wave analyser or like instrument will yield the amplitude spectrum. The phase spectrum may be measured in an analogous manner by a modification of such an instrument.

The input to the system to be optimized is assumed to consist of some signal contaminated with noise. A particular frequency component of the signal in the time-plane

may be written as the real part of a vector rotating in the complex plane at that angular frequency, ω, and having for its magnitude the value of the signal amplitude spectrum $\mathcal{S}(\omega)$. It is denoted by

$$\mathrm{Re}\left[\mathcal{S}(\omega)\,\epsilon^{j\omega t}\right]. \tag{1}$$

The ω component of the noise may be represented similarly, with the addition of a relative phase angle, $\phi(t;\omega)$, which refers the phase angle of the ω component of the noise voltage, in the time-plane, to that of the signal voltage. The nature of the relative phase angle as a function of time will be discussed later in this derivation. The ω component of the noise input is denoted by

$$\mathrm{Re}\left[\mathcal{N}(\omega)\,\epsilon^{j\omega t + j\phi(t;\omega)}\right]. \tag{2}$$

The sum of the signal and noise vectors represents the ω component of the system input, consisting of signal contaminated with noise. It is written as

$$\mathrm{Re}\left[\mathcal{S}(\omega)\,\epsilon^{j\omega t} + \mathcal{N}(\omega)\,\epsilon^{j\omega t + j\phi(t;\omega)}\right]. \tag{3}$$

The system output spectral density is the result of the above input operating on the system transfer function, denoted by the familiar $M(\omega)\,\epsilon^{j\alpha(\omega)}$, where $M(\omega)$ is the closed loop gain and $\alpha(\omega)$ is the phase shift. A particular frequency component of this output spectrum is written as

$$\mathrm{Re}\left[\left\{\mathcal{S}(\omega)\,\epsilon^{j\omega t} + \mathcal{N}(\omega)\,\epsilon^{j\omega t + j\phi(t;\omega)}\right\} M(\omega)\,\epsilon^{j\alpha(\omega)}\right]. \tag{4}$$

A frequency component of the error is the difference between the signal and the output components, expressions (1) and (4) above. The real part of this frequency component of

the error is

$$e(t;\omega) = \mathcal{S}(\omega)\,\mathrm{Cos}(\omega t) - \mathcal{S}(\omega)\,M(\omega)\,\mathrm{Cos}(\omega t+\alpha) - \mathcal{N}(\omega)\,M(\omega)\,\mathrm{Cos}(\omega t+\alpha)\,\mathrm{Cos}(\phi(t;\omega)) + \mathcal{N}(\omega)\,M(\omega)\,\mathrm{Sin}(\omega t+\alpha)\,\mathrm{Sin}(\phi(t;\omega)). \quad (5)$$

In order to form the particular frequency component of the mean square error spectrum, the error of that component, equation (5), must be squared and averaged over its indicated variations as a function of time. The magnitudes of the signal and noise amplitude spectra and the system function gain and phase shift are held temporarily constant. The instantaneous amplitude of each component of the signal and noise inputs was defined to vary in time as described by the real part of the rotating, complex vector. However, no definite variation in time was assumed for the relative phase angle, $\phi(t;\omega)$. If the signal and noise are completely correlated, this angle will have a constant value. For zero correlation between signal and noise, the variation of the angle will be uniformly random, taking on all values between zero and 2π with equal probabilities. For intermediate degrees of correlation, the angle will vary slowly about some average value. Because the nature of the variation in time of the relative phase angle is markedly different from the cyclic variation in time of the instantaneous amplitudes, these two functional forms may be assumed to have no cross-correlation in their variation and may be averaged independently. The averaging of the cyclic variations of the squared error spectrum is accomplished by

integration over one cycle. The time-average of the cosine and sine of the relative phase angle will be denoted by the conventional bar over the expression, e.g., $\overline{\cos(\phi(t;\omega))}$. The result of averaging the square of equation (5) is thus

$$\overline{e^2}(\omega) = \tfrac{1}{2}\left\{ \delta^2(\omega) + (\delta^2(\omega) + \eta^2(\omega))M^2(\omega) - 2\,\delta^2(\omega)\,M(\omega)\,\cos(\alpha(\omega)) + 2\,\delta(\omega)\,\eta(\omega)\,M(\omega)\left[M(\omega) - \cos(\alpha(\omega))\right]\overline{\cos(\phi(t;\omega))} + 2\,\delta(\omega)\,\eta(\omega)\,M(\omega)\,\sin(\alpha(\omega))\,\overline{\sin(\phi(t;\omega))}\right\}. \quad (6)$$

For the commonly assumed case of zero correlation between signal and noise, the variation in time of the relative phase angle was noted above to be uniformly random. The averages of its sine and cosine values are in that case zero, and the mean square error spectrum is accordingly simplified. For other values of correlation between signal and noise, the averages of the sine and cosine of the relative phase angle are not intuitively known. However, they may be determined by either of two ways, depending upon the form of the information available. If the probability density function which describes the random behavior of the angle is known, the average values may be calculated by use of the Expected Value operation, explained in Part II of this dissertation. If the values must be determined experimentally, the angle may be measured by a modification of two identical wave analysers or like instruments, using a common oscillator and having a rapidly

responding phase detector coupled between them. The sine and cosine of the angle may be determined from the phase detector output by a suitable resolver, and their averages may be read on two appropriately averaging voltmeters.

Because $\mathcal{S}(\omega)$ and $\mathcal{N}(\omega)$ were considered to be the magnitudes of the input signal and noise amplitude spectra, they are peak values. However, amplitude spectra are commonly given or measured in rms values, as would be obtained by taking the square root of the power spectral densities. Therefore the symbols $S(\omega)$ and $N(\omega)$ will henceforth be used to represent the rms values of the signal and noise amplitude spectra, and the factor of 1/2 which multiplies equation (6) is removed. The mean square error spectrum may be written for brevity as

$$\overline{e^2}(\omega) = S^2 + (S^2+N^2)M^2 - 2S^2M\,\mathrm{Cos}(\alpha) + 2SNM(M-\mathrm{Cos}(\alpha))\,\overline{\mathrm{Cos}(\phi)} + 2SNM\,\mathrm{Sin}(\alpha)\,\overline{\mathrm{Sin}(\phi)}\,. \tag{7}$$

By proper interpretation and substitution of analogous symbols (1c,3), this spectrum may be shown to be equivalent to that commonly found in the literature. The prospect of easier application of the mean square error spectrum derived above to many experimental studies has been discussed earlier.

PART II

THE OPTIMIZATION PROBLEM

The Method of Solution

The mean square error spectrum is a function of the signal and noise amplitude spectra, the averaged sine and cosine of the relative phase angle, and the system gain and phase shift. If any exhibit some degree of random time-variation, the mean square error spectrum must be averaged over all of their possible values to form the expected or most probable value of that spectrum. It is this expected mean square error spectrum that represents the over-all performance of the system and is to be minimized by proper choice of the system gain and phase shift. This process of averaging the spectrum over all random values of the input spectra and system function is accomplished by the use of a probability theory operation termed Expected Value (4a) or Mathematical Expectation (5a). Appendix A of this dissertation contains a description of an experimental check on this use of the Expected Value operation.

A limitation on the rate of random time-variation results from the separation of the averaging process into two steps. The first is the time average of each frequency component of the squared error spectrum over one cycle of its angular frequency. The second step is the Expected

Value operation, averaging each frequency component of the mean square error spectrum over all random variations in the input spectra and system parameters. Therefore the random variation must be slow enough, compared to the angular frequency of the component being averaged, that the separation of the averaging process is valid.
Since the kind of random time-variations referred to here generally occur very slowly, or at infrequent intervals, this limitation may be assumed to be met.

The Expected Value of a function of any number of random variables is defined (4a) as the average value of that function when its random variables take on all of their permissible values with their associated probabilities of occurrence. Briefly, the Expected Value of a function, $f(x)$, of the random variable, X , where X takes on each value with a probability $p(x)$, is denoted by $E[f(x)]$. Either it is determined mathematically by a) summation: $E[f(x)] = \sum_{\text{all } x} f(x) \cdot p(x)$, if X takes on only discrete values in which case $p(x)$ is termed the discrete probability density function of X ; or it is determined by b) integration: $E[f(x)] = \int_{\text{all } x} f(x) \cdot p(x) \cdot dx$, if X varies continuously in which case $p(x)$ is termed the probability density function of X . For all functions of any practical importance (5a), the following properties of the Expected Value operation are valid. The Expected Value of a sum is the sum of the Expected Values. The Expected Value of a constant is that

constant. The Expected Value of a product of functions of statistically independent random variables is the product of their separate Expected Values. The Expected Value of a function, $f(x,y,z)$, of several random variables, x , y , and z , having a joint probability density function, $p(x,y,z)$, is

$$E[f(x,y,z)] = \iiint_{\text{all } x,y,z} f(x,y,z)\ p(x,y,z)\ dx\ dy\ dz\ .$$

The requirements for a probability density function, $p(x)$, are that $p(x) \geq 0$ and that $\int_{\text{all } x} p(x)\,dx = 1$ since it is certain that x will take on some value within its range of values.

As mentioned above, any or all of the quantities composing the mean square error spectrum may exhibit random time-variation. In addition, the variations of any quantity may be dependent in the probability sense on the variations of any or all of the other quantities. Because of the many possibilities for such varied and interrelated random behavior, the method of solution will be presented in the form of several cases. For each case, a different set of restrictions on the many possibilities for dependent and independent random behavior will be used. These cases are divided into two groups. The division results from two possible procedures in using the mean square error spectrum, as is explained in the next paragraph.

The first procedure involves minimization of each frequency component (or ordinate) of the mean square error spectrum. There results the optimum system transfer function in terms of the input spectra. The second procedure requires

minimization of the total mean square error, which has been shown in the literature to be the integral of the mean square error spectrum over all real, positive frequencies. Each procedure has its special advantages and disadvantages. The first procedure yields generalized results in a relatively straightforward manner. The second procedure permits a generalized formulation of its solution. However, to achieve useful results the integration over all frequency requires the substitution of the type of system transfer function to be optimized. Specific signal and noise amplitude spectra and the averages of the sine and cosine of the relative phase angle must also be substituted. These useful results are the optimum values of the parameters of the type of system substituted, given in terms of the coefficients of the spectra chosen. Thus the second procedure assures physically realizable, optimum systems. However, the first procedure will be shown to specify that the optimum system has a gain characteristic determined by the shapes and relative magnitudes of the input spectra and a phase shift characteristic everywhere zero. Such a system is seldom physically realizable because it is generally impossible to have a rate of change of attenuation with respect to frequency without having non-zero phase shift. In fact, for minimum phase shift, stable, linear systems composed of fixed, lumped elements, Bode's theorms (2a) and others indicate that the phase shift at any frequency is a unique function of the attenuation at all

frequencies. In spite of this limitation the generality of the first procedure yields valuable qualitative insight into the synthesis problem. In addition, by yielding the optimum though nonrealizable system transfer function, it may be used to calculate the lower bound of the mean square error for any particular input spectra. This will be a limit which the mean square error of physically realizable systems may approach.

Recognizing the importance of both of the procedures described above, this dissertation will treat the procedures separately in the two major sections that follow. These are "Optimum, Idealized System Transfer Functions" and "Optimization of Physically Realizable Systems". As mentioned above, each section will include several cases which illustrate the solution for problems characterized by various types of random time-variation.

Optimum, Idealized System Transfer Functions

Independently Time-varying Input Spectra. For the solution of a case which illustrates both the first procedure for using the mean square error spectrum and the treatment for random time-variations, the following assumptions and restrictions are made. The signal and noise amplitude spectra are assumed to be independently time-varying, having probability density functions $p(S(\omega))$ and $q(N(\omega))$. Each density function describes the probability of occurrence of each amplitude that a particular frequency component (ω) of the time-varying spectrum exhibits for the set of measurements of the spectrum.

These probability density functions may be any one of the familiar: Uniform, Gaussian or Normal, Gamma, Beta, or they may be some undefined density function. As will be shown, the shape of the density function need not be known; only its first and second moments are used in mean square minimization. The assumption of independence between signal and noise implies the restriction that their relative phase angle has a uniform probability density, taking on all values from zero to 2π with equal probabilities. Thus the averages of the sine and cosine of $\phi(t;\omega)$ in the mean square error spectrum of equation (7), are zero. The system function is restricted to be time-invariant for the case following. The Expected Value of the mean square error spectrum, equation (7), under these assumptions, is

$$E[\overline{e}^2(\omega)] = E[S^2] + M^2(E[S^2]+E[N^2]) - 2M\,Cos(\alpha)\,E[S^2]. \quad (8)$$

The various Expected Values may be evaluated as follows:

$$E[S] = \int_{all\ S} S\cdot p(S)\cdot dS = a_1 \ , \quad E[S^2] = \int_{all\ S} S^2\cdot p(S)\cdot dS = a_2 \ ,$$
$$E[N] = \int_{all\ N} N\cdot q(N)\cdot dN = n_1 \ , \quad E[N^2] = \int_{all\ N} N^2\cdot q(N)\cdot dN = n_2 \ ,$$

where a_1 and n_1 are the averages or first moments, and a_2 and n_2 are the second moments. The second moment is often written in terms of both the first moment and the variance. The variance, denoted by σ^2, is the second moment about the mean, and is a measure of the spread of a probability density function about its mean value. The positive square root of variance, σ, is commonly termed the standard devia-

tion. For a random variable x with a mean value X_1 , the variance of x is

$$\sigma_x^2 = E[(x-X_1)^2] = E[x^2] - 2X_1 E[x] + X_1^2 = X_2 - X_1^2.$$

Therefore: $E[S^2] = A_1^2 + \sigma_S^2$, and $E[N^2] = n_1^2 + \sigma_N^2$.

These moments are functions of frequency, being spectra describing the sets of signal and noise amplitude spectra. Substituting these evaluations in the expected mean square error spectrum yields $E[\overline{e^2}(\omega)] = (A_1^2 + \sigma_S^2) +$

$$M^2(A_1^2 + \sigma_S^2 + n_1^2 + \sigma_N^2) - 2(A_1^2 + \sigma_S^2) M \cos(\alpha). \quad (9)$$

These moments are of special interest because they indicate the information which is necessary for the use of the mean square criterion. The first moment is the average of the amplitude spectra which describe the time-varying spectra. The second moment is the average of the squares of the amplitude spectra (the mean square); it is thus the average of the power spectra describing the signal or noise time-varying spectrum (the first moment of the power spectra). The first and second moments of the amplitude spectra constitute all the information about the time-variation in input spectra or system function that the mean square criterion requires. No other information can be used. For this case of uncorrelated signal and noise, the method requires only the average of the power spectra; the first moments of the amplitude spectra are not required. The averages referred to above are all weighted according to the probabilities of occurrence of the various spectra, as

the Expected Value operation indicates. Writing the second moment of the amplitude spectra as the square of their first moment plus their variance permits using the variance term to illustrate the effect of random time-variations in the input spectra.

The synthesis problem in this case is the determination of the system gain and phase shift as functions of the signal and noise input spectra to minimize the expected mean square error spectrum. The phase shift (α) at a particular frequency is a unique function of the gain (M) at all frequencies for nearly all networks and systems (ie. for minimum phase shift, linear, stable systems composed of fixed, lumped elements), as expressed by Bode (2a) and others. If this function could be substituted in the error spectrum, the derivative with respect to M could be equated to zero to determine the optimum M. From this the optimum α would be calculated. Since the relationship between M and α requires treatment of all frequencies concurrently, and is so intractable as to prevent even that use of it, the classical assumption of α independent of M will be made. Therefore a much less restricted system function will be optimized. It includes all physically realizable systems as a special case. The resulting gain and phase shift characteristics will be the best possible although probably not physically realizable. The assumption of independence between α and M allows the partial derivatives of any component of the

spectrum, taken with respect to α and M, to be each equated to zero and solved for their optimum values.

$$\partial E[\overline{e^2}(\omega)]/\partial\alpha = 2 A_2 \sin(\alpha) = 0, \qquad (10)$$

which requires that $\alpha=0$; that is, a zero phase shift system is indicated to be optimum.

$$\partial E[\overline{e^2}(\omega)]/\partial M = 2M(A_2 + n_2) - 2A_2\cos(\alpha) = 0$$

yields, with $\alpha=0$,

$$M(\omega) = \frac{A_2(\omega)}{A_2(\omega) + n_2(\omega)}, \qquad (11)$$

If the second moments are denoted by P_S and P_N, the averages of the sets of signal and noise power spectra, then

$$M(\omega) = \frac{P_S}{P_S + P_N}$$

which result has been developed by Wiener (6) in a basically different manner.

Writing the second moment as the square of the first moment plus the variance permits using the variance term to illustrate the effect of random time-variations in the input spectra. From equation (11), the gain of the optimum, zero phase shift system transfer function is

$$M(\omega) = \frac{A_1^2 + \sigma_S^2}{A_1^2 + \sigma_S^2 + n_1^2 + \sigma_N^2} = \frac{1}{1 + (n_1^2 + \sigma_N^2)/(A_1^2 + \sigma_S^2)}.$$

The latter form indicates that the effect of time-varying behavior in a particular frequency component of the signal spectrum is to increase the optimum value of closed loop system gain toward unity at that frequency. The effect of time-varying behavior in the noise spectrum is conversely to decrease the closed loop gain toward zero. It is interesting to note that the zero frequency closed loop gain of

the system should not necessarily be unity as is generally seen in the literature. Consideration of the expression

$$M(0) = 1 / (1 + \eta_2(0)/\Phi_2(0))$$

yields the conclusion that the optimum zero frequency gain depends upon the limit of $\eta_2(\omega)/\Phi_2(\omega)$ as ω approaches zero. That gain will lie between unity and zero, the end points being special cases of the noise or signal spectra alone being zero at zero frequency. Of course the analogous conclusion is valid at any frequency. Note that an optimum value of closed loop gain, M, greater than unity is never indicated.

The characteristic of an optimized, physically realizable system transfer function will tend toward the non-realizable, idealized characteristic specified in this section. However, the realizable design will necessarily involve compromises. It must smooth out any sharp peaks or valleys in the idealized gain characteristic as a result of the inherent phase shift and fixed shape of any type of realizable system transfer function.

The next two cases will constitute a sufficient illustration of the method as applied in this first section of Part II of the dissertation. In the next case the signal and noise spectra will be assumed to be dependently time-varying .In the final case the system function will be assumed to exhibit some independent, random, time-varying behavior, and the input spectra will be assumed to be time-invariant.

Dependently Time-varying Input Spectra. In this the second case of Part II, the signal and noise amplitude spectra are assumed to be dependently time-varying, having therefore a joint probability density function $f(S,N)$. The average values of the sine and cosine of the relative phase angle are also assumed to vary in time, independently of the amplitudes for simplicity. The probability density for the averaged sine is denoted by $p(\overline{\sin\phi})$. It is defined for $-1 \leq \overline{\sin\phi} \leq +1$ and is assumed to be an even function about zero with its most probable value at zero also. The probability density for the averaged cosine is similarly denoted by $q(\overline{\cos\phi})$. It is assumed to have its most probable value at plus one. The system function is to remain time-invariant for this case. The form of the expected mean square error spectrum under these assumptions is

$$\begin{aligned} E[\overline{e^2}(\omega)] = {} & E[S^2] + (E[S^2]+E[N^2])M^2 - \\ & 2\,E[S^2]\,M\cos(\alpha) + \\ & 2\,E[SN]\,M\,(M-\cos(\alpha))\,E[\overline{\cos\phi}] + \\ & 2\,E[SN]\,M\sin(\alpha)\,E[\overline{\sin\phi}]. \end{aligned} \qquad (12)$$

The Expected Values of S^2 and N^2 are determined in the same manner as was given in the preceding case.

$$E[S^2] = \mathcal{A}_2 = \mathcal{A}_1^2 + \sigma_S^2 \quad , \quad E[N^2] = \eta_2 = \eta_1^2 + \sigma_N^2 \, ,$$

where $\mathcal{A}_1$ and η_1 are the first moments, $\mathcal{A}_2$ and η_2 are the second moments, and σ_S^2 and σ_N^2 are the variances.

An evaluation which did not appear in the previous case is

$$E[SN] = \iint_{\text{all } S,N} S\,N\cdot f(S,N)\cdot dS\,dN = \rho\,\sigma_S\,\sigma_N + \mathcal{A}_1\eta_1$$

where ρ is termed the correlation between S and N.

If ρ is zero, S and N are said to be independent and the above Expected Value would be obtained simply by

$$E[SN] = E[S] \cdot E[N] = A_1 n_1 .$$

The Expected Values of the averaged sine and cosine are

$$E[\overline{\sin\phi}] = \int_{-1}^{+1} \overline{\sin\phi} \cdot p(\overline{\sin\phi}) \cdot d(\overline{\sin\phi}) = 0$$

because of the assumed nature of $p(\overline{\sin\phi})$, and

$$E[\overline{\cos\phi}] = \int_{-1}^{+1} \overline{\cos\phi} \cdot q(\overline{\cos\phi}) \cdot d(\overline{\cos\phi}),$$

which is assumed to have some value between zero and unity, depending upon the shape of $q(\overline{\cos\phi})$. Let that value be denoted by P. Substitution in the expected mean square error spectrum, equation (12), yields

$$E[\overline{e^2}(\omega)] = (A_1^2 + \sigma_S^2) + (A_1^2 + \sigma_S^2 + n_1^2 + \sigma_N^2) M^2 - 2(A_1^2 + \sigma_S^2) M \cos(\alpha) + 2(A_1 n_1 + \rho \sigma_S \sigma_N) M(M - \cos(\alpha)) P.$$

Partial differentiation as before yields

$$\partial E[\overline{e^2}(\omega)]/\partial\alpha = 2(A_1^2 + \sigma_S^2) \cdot M \cdot \sin(\alpha) + 2(A_1 n_1 + \rho \sigma_S \sigma_N) \cdot M \cdot \sin(\alpha) \cdot P = 0.$$

This requires that either

$$(A_1^2 + \sigma_S^2) + P \cdot (A_1 n_1 + \rho \sigma_S \sigma_N) = 0$$

or $\sin(\alpha) = 0$ and therefore $\alpha = 0$ at all frequencies. Setting the partial derivative with respect to M equal to zero yields

$$M(\omega) = \frac{(A_1^2 + \sigma_S^2) + (A_1 n_1 + \rho \sigma_S \sigma_N) \cdot P}{(A_1^2 + \sigma_S^2) + (n_1^2 + \sigma_N^2) + 2(A_1 n_1 + \rho \sigma_S \sigma_N) \cdot P} \cos(\alpha) \quad (13)$$

The first condition on α makes the closed loop gain equal zero. Therefore the mean square error would be as large as if there were no system. The other condition, $\alpha = 0$,

assures a minimum of the error with respect to α for

$$(a_1^2 + \sigma_s^2) + (a_1 m_1 + \rho \sigma_s \sigma_N) P > 0.$$

Although ρ may range between minus and plus one, the inequality will generally be satisfied. For this condition, the optimum system has zero phase shift everywhere and the gain characteristic, $M(\omega)$, given by equation (13) with $\alpha = 0$.

Time-varying System Closed Loop Gain. In the final case of this section, the system transfer function is assumed to exhibit some independent, random, time-varying behavior. The phase shift is assumed to be everywhere zero. The gain is assumed to have a probability density function $p(M)$. The signal and noise are assumed to be uncorrelated and their amplitude spectra are time-invariant. The expected mean square error spectrum is therefore

$$E[\overline{e^2}(\omega)] = S^2 + (S^2+N^2)E[M^2] - 2S^2 E[M].$$

The required Expected Values are

$$E[M] = \int_{\text{all } M} M \cdot p(M) \cdot dM = m_1,$$
$$E[M^2] = \int_{\text{all } M} M^2 \cdot p(M) \cdot dM = m_2 = m_1^2 + \sigma_m^2.$$

The optimization problem is to determine the best value of m_1, the average of the randomly varying system closed loop gain, $M(\omega)$. The variance of that random variation is assumed to be a fixed fraction of the average. Therefore

$$m_1^2 + \sigma_m^2 = m_1^2 + m_1^2 \delta^2 = m_1^2 (1+\delta^2).$$

This is more reasonable than to assume a variance which is

independent of the average, whether that average is large or small. Substitution in the expected mean square error spectrum yields

$$E[\bar{e}^2(\omega)] = S^2 + (S^2+N^2)\,m_1^2(1+\delta^2) - 2S^2 m_1 . \qquad (14)$$

Minimization of this spectrum with respect to m_1 by setting its derivative equal to zero yields

$$m_1 = \frac{S^2}{(S^2+N^2)(1+\delta^2)} . \qquad (15)$$

A conclusion evident here is that random variation in the closed loop system gain dictates a smaller value of average gain that would otherwise be designed. Another conclusion is of more importance; it may be shown that if the randomness of the gain were neglected, a larger value of mean square error would always result. To demonstrate this conclusion, a ratio of two evaluations of the expected mean square error, equation (14) above, is formed. The ratio is the evaluation using the value of m_1 derived above (equation (15)) by properly treating the randomness of the gain, divided by the evaluation using $m_1 = \frac{S^2}{S^2+N^2}$ which results from treating the gain as a nonrandom parameter. The ratio is

$$\frac{1}{1 + \frac{\delta^4 S^2}{S^2\delta^2 + N^2(1+\delta^2)}} .$$

Thus the conclusion immediately above is demonstrated by this ratio being always less than unity, except in the limit of S or δ equal to zero. If $S=0$, there is no signal input and both values of gain are zero, hence both have the same error. If $\delta=0$, there is no randomness and both values of

gain are the same, hence the errors are the same.

Optimization of Physically Realizable Systems

Thus far, application of this method has been restricted to systems having zero phase shift. Recognizing the general impossibility of a rate of change of attenuation without phase shift in physically realizable systems, the second procedure described previously will be considered in detail in this section. The most advantageous use of this more realistic procedure lies in substitution of a particular type of system transfer function for the general expression, $M\epsilon^{j\alpha}$, in the mean square error spectrum. Particular forms of the signal and noise amplitude spectra are also substituted. The procedure involves integrating the mean square error spectrum over all real, positive frequencies to form the mean square error. For problems characterized by random time-variations in the input spectra and system transfer function, the average of the mean square error over those variations must be determined. This is accomplished by the use of the Expected Value operation as described and applied earlier. The problem of system optimization is that of minimizing the expected mean square error with respect to such system parameters as amplification and time constants, within the limitations of system stability. Such minimization is carried out by ordinary or partial differentiation or by graphical means. There results the optimum values of the system parameters and the optimum

characteristic of the type of system transfer function substituted. The conclusions for the case of independent, random time-variations in the input spectra are much the same as given for the first procedure. These are: randomness in the signal spectrum dictates an increase in system gain for optimization, while randomness in the noise spectrum dictates a decrease in system gain. The second moments of the amplitude spectra as described before will dictate the proper values of system parameters. The conclusions for cases of randomness in the system transfer function are determined by this second procedure in a different manner. In contrast to simply attributing the randomness to the entire closed loop gain function as was necessary under the first procedure, this second procedure permits ascribing the randomness directly to one or more specific system parameters. Examples of such random variation are drifting amplifications and fluctuating time constants. As a result of this contrast, the cases of randor time-variation in this section will be restricted to randomness in system time constants.

The general solution to the optimization problem as described above is formulated below in mathematical terms, illustrating this second procedure symbolically. Particular solutions for some restricted cases are obtained later in this section. For the general formulation, the input signal and noise amplitude spectra are denoted by products of

power series in ω:

$$S(\omega) = \sum_{k=0}^{K} \beta_k \omega^k \Big/ \sum_{\nu=0}^{K+1} \delta_\nu \omega^\nu \quad , \quad N(\omega) = \sum_{k=0}^{L} \eta_k \omega^k \Big/ \sum_{\nu=0}^{L+1} \xi_\nu \omega^\nu .$$

These input spectra operate on a system transfer function given in general by:

$$M\epsilon^{j\alpha} = \prod_{k=1}^{N} (\rho_k + j\omega\gamma_k) \Big/ \prod_{\nu=1}^{N+1} (\lambda_\nu + j\omega\tau_\nu), \text{ for which}$$

$$M(\omega) = \prod_{k=1}^{N} \sqrt{\rho_k^2 + \omega^2\gamma_k^2} \Big/ \prod_{\nu=1}^{N+1} \sqrt{\lambda_\nu^2 + \omega^2\tau_\nu^2} \quad \text{and}$$

$$\alpha(\omega) = \sum_{k=1}^{N} \tan^{-1}\left(\frac{\gamma_k\omega}{\rho_k}\right) - \sum_{\nu=1}^{N+1} \tan^{-1}\left(\frac{\tau_\nu\omega}{\lambda_\nu}\right).$$

The general mean square error spectrum, $\overline{e^2}(\omega)$, is the result of substitution of the terms above in equation (7). The formulation of the general solution is a set of simultaneous equations

$$\frac{\partial}{\partial q_p} E\left[\int_0^\infty \overline{e^2}(\omega)\cdot d\omega\right] = 0,$$

where q_p is successively each of the system parameters (ρ_k, γ_k, λ_ν, τ_ν) which are considered to be subject to adjustment. The resulting values of system parameters must satisfy the restriction that

$$\prod_{\nu=1}^{N+1} (\lambda_\nu + j\omega\tau_\nu) = 0$$

have no roots with positive real parts, thus insuring system stability. The zero of each partial derivative must be examined to insure that it represents a minimum of mean square error ather than a maximum. The Expected Value operation is carried out over all system function and input spectra parameters. The random behavior of these parameters is described

in general by a joint probability density function

$$p(\beta_k, \delta_\nu, \eta_k, \xi_\nu, \rho_k, \gamma_k, \lambda_\nu, \tau_\nu).$$

An alternate and more compact formulation may be developed from the concept of a system space as

$$\text{grad } E\left[\int_0^\infty \overline{e^2}(\omega)\cdot d\omega\right] = 0.$$

The formulation above has been continued to completion for some specific classes of problems. These solutions are described below. The general solution has not been treated further in this dissertation because further general treatment is not feasible.

For the first class of problems below, this procedure is applied to a set of time-invariant servosystems. The approach in that section is somewhat more general than that found in the literature (1b). A discussion of the solutions illustrates some of the conclusions resulting from the use of the mean square criterion. The invariant class is primarily included as an introduction to the later classes, which are characterized by random, time-varying system time constants.

Optimization of Invariant Systems. There is described in this part of the paper the application of the procedure just formulated to three general types of servosystems. The signal amplitude spectrum is assumed to be of first order form, as shown in figure 1. A white noise spectrum is assumed. The correlation between signal and noise is assumed to be zero. The signal to noise ratio is described

Input Amplitude Spectra

Signal: $S(\omega) = \frac{\beta}{\sqrt{\omega^2 + \delta^2}}$

Noise: $N(\omega) = \eta$

Signal to Noise Ratio: $\Sigma = \frac{\beta}{\delta \eta}$

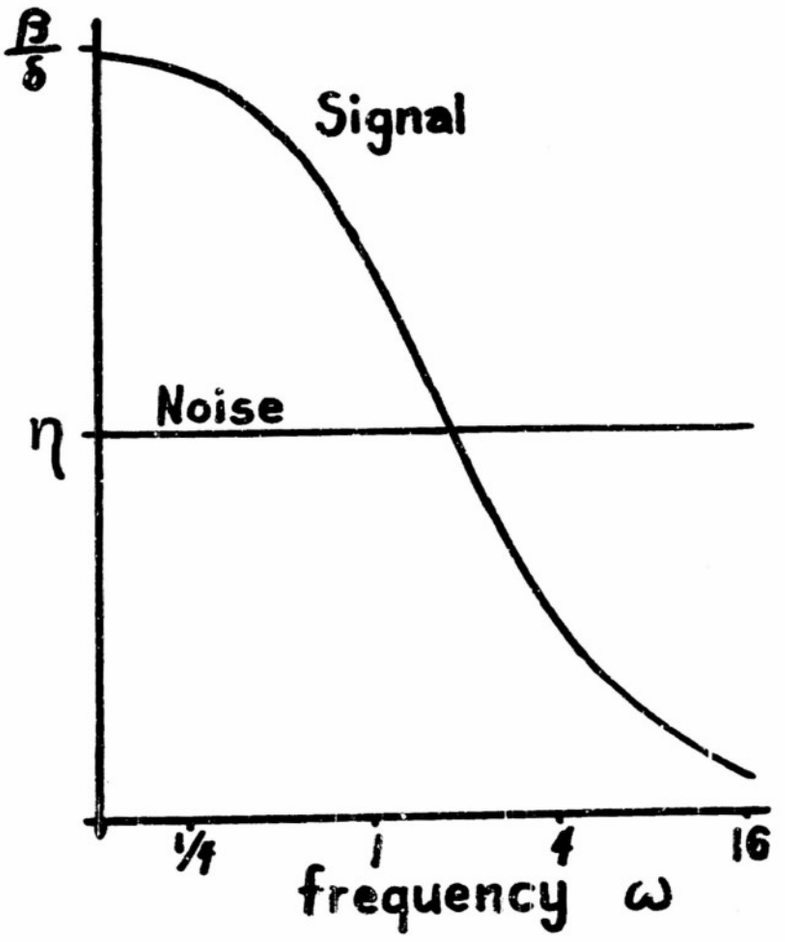


Figure 1

in terms of the values of the signal and noise amplitude spectra at zero frequency (dc). The ratio is denoted by Σ and is an important parameter in the examples to follow. The series of general servosystems to be optimized may be described by their open loop transfer functions, which are (figure 2): (1) an integrator, (2) a simple lag, and (3) a simple lag with an ideal amplifier.

For each example of the series, the input spectra shown in figure 1 and the system gain and phase shift are substituted in the proper form of the mean square error spectrum.

The spectrum is integrated over all real frequencies, by the method of residues, to form the mean square error. The values of the system parameters which minimize the mean square error for a given ratio of signal to noise are determined by ordinary or partial differentiation. These values are subject to the requirements of system stability. The parameters of these three systems are, for system (1) the integration constant τ, (2) the time constant τ, and (3) the gain K and time constant τ. Ordinary differentiation is used to optimize the first two, single-parameter systems. In more complex examples of two-parameter systems, the partial differentiation becomes too intractable, and the surface generated by the mean square error as a function of the two parameters is better examined graphically for its minimum. The mathematics involved in this applica-

Servosystems

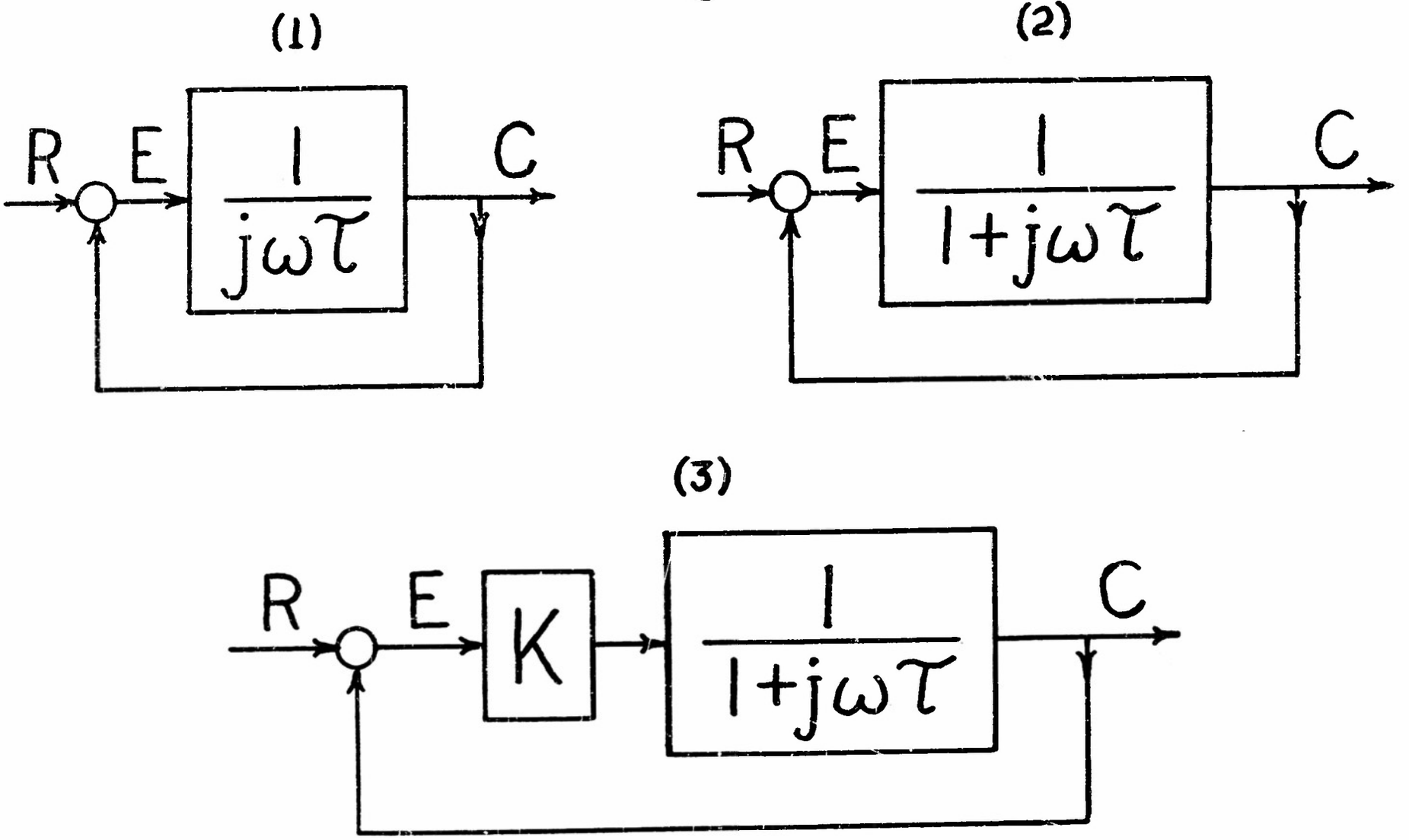


Figure 2

tion is straightforward in essence.

As a more detailed illustration of the method, the essential points in the optimization of the first system are presented here. The signal and noise amplitude spectra are given by

$$S(\omega) = \frac{\beta}{\sqrt{\omega^2 + \delta^2}} \quad , \quad N(\omega) = \eta \, .$$

The closed loop system transfer function of the integrator is

$$M e^{j\alpha} = \frac{1}{1 + j\omega\tau} \quad , \quad \therefore \; M = \frac{1}{\sqrt{1 + \omega^2\tau^2}} \, ,$$

$$\alpha = -\tan^{-1}(\omega\tau) = \cos^{-1}\left(\frac{1}{\sqrt{1 + \omega^2\tau^2}}\right) .$$

These quantities are substituted in the mean square error spectrum, equation (7), which is written here for zero correlation between signal and noise.

$$\overline{e^2}(\omega) = S^2 + (S^2 + N^2)M^2 - 2S^2 M \cos(\alpha) .$$

Substitution yields

$$\overline{e^2}(\omega) = \frac{\beta^2}{\omega^2 + \delta^2} - \frac{\beta^2}{\omega^2 + \delta^2} \cdot \frac{1}{1 + \omega^2\tau^2} + \frac{\eta^2}{1 + \omega^2\tau^2} \, .$$

Integration by residues results in the mean square error,

$$\overline{e^2} = \frac{\pi}{2}\left\{\frac{\beta^2}{\delta} - \frac{\beta^2}{\delta(1 + \delta\tau)} + \frac{\eta^2}{\tau}\right\},$$

which may be written in terms of the signal to noise ratio,

$$\Sigma = \beta/\delta\eta \, , \text{ as } \quad \overline{e^2} = \frac{\pi}{2}\frac{\beta^2}{\delta}\left\{1 - \frac{1}{1 + \delta\tau} + \frac{1}{\Sigma^2\delta\tau}\right\}. \qquad (17)$$

Minimization of this expression is accomplished by setting the derivative of $\overline{e^2}$ with respect to τ equal to zero and solving for the optimum τ, within the system stability requirement that τ be nonnegative. This yields

$$\delta\tau = \frac{-1}{\Sigma + 1} \, , \; \frac{+1}{\Sigma - 1} \, .$$ The first root yields a maximum of $\overline{e^2}$ and an unstable system. The second root yields

a minimum of $\overline{e^2}$ and hence $\delta\tau = \frac{1}{\Sigma - 1}$ is the optimum time constant. This result will be discussed below.

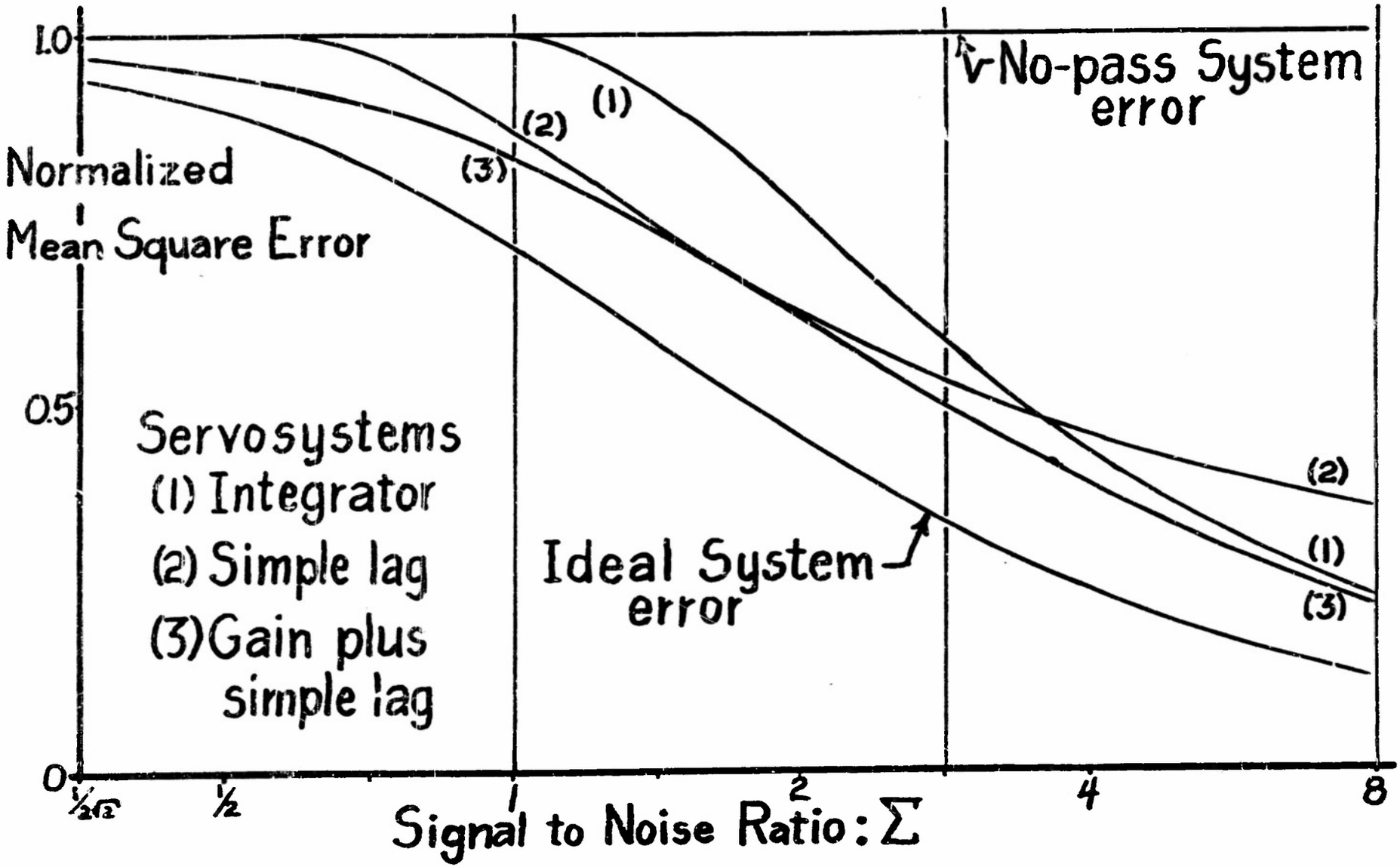
The results of optimizing the three types of servo-systems listed above are presented in figure 3. The mean square error of each optimized system is shown as a function of the signal to noise ratio Σ. The mean square error function for the zero phase shift system previously derived, equation (11), is included for comparison, despite the unrealizability of that system, because it represents a lower bound on the mean square error attainable. These functions are normalized to the mean square error of a no-pass system (a system having no output). Figure 3 illustrates several important conclusions of interest for both system analysis and synthesis. These are:

(1) A worthwhile integrator system cannot be designed for $\Sigma < 1$; all stable integrator systems have at least as much error as the no-pass system for any such sigma. The reason is that the integration constant yielding minimum mean square error, given by $\tau = \frac{1}{\delta(\Sigma - 1)}$, becomes infinite as Σ approaches unity, that is, the system bandwidth becomes zero. For $\Sigma < 1$, the optimum τ becomes negative, hence the system is unstable, and any positive τ yields more error than a no-pass system.

(2) The mean square error curve of the simple lag system is similar to that of the integrator. The optimum time constant is given by $\tau = \frac{2}{\delta(\sqrt{3}\Sigma - 1)}$ which becomes in-

Figure 3

System Mean Square Errors



finite when $\Sigma = 1/\sqrt{3}$. For smaller Σ the system is worthless. The system has a gain of 1/2 at zero frequency. That this is desirable for Σ near unity is shown by its superiority to the integrator system which has a gain of unity at zero frequency. For Σ larger than 3.7 the integrator system is seen to be superior, indicating the desirability of unity gain at zero frequency for large signal to noise ratios.

(3) The simple lag with an ideal amplifier is seen to be superior to the first two systems almost everywhere. This superiority may be attributed to the gain parameter which permits changing the zero frequency gain, and hence the entire level of system gain, according to the ratio of signal to noise. At the point where $\Sigma = \sqrt{3}$, this system coincides with that of system (2) because its optimum gain is unity. Below $\Sigma = \sqrt{3}$, its optimum gain is less than unity; the amplifier is only an attenuator. In this region of $\Sigma < \sqrt{3}$, system (2) could be improved easily by the addition of an attenuator in its open loop, making it the same as system (3).

(4) In addition to the systems described, a second order system was similarly treated. It has a closed loop transfer function composed of the product of two simple lag circuits with independent time constants. Optimization of this system with respect to the two time constants dictated that one of them be zero for minimum mean square error. Thus, under the assumptions made, the best second order

system of this form is a simple first order system corresponding to the integrator system treated first.

The optimization procedure for the two-parameter systems discussed above as (3) and (4) is described in detail in Appendices B and C of this dissertation.

Optimization of Random System Parameters. In the preceding class of problems, the servosystem was considered to be composed of fixed elements, that is of elements such as gain and time constants having constant values during the period of operation for which the system was being designed. It is possible that the environment of the system may vary greatly during that period. As a result of changes in pressure, temperature, humidity, and the like, the gain and time constants may vary appreciably from their intended values. Such variations are of more or less random nature.

For this case the system gain (M) and phase shift (α) are assumed to exhibit some random behavior independent of the signal and noise. The signal and noise spectra are assumed to be constant and to have zero correlation over the period of operation. However, the method could handle cases of dependence in the probability sense between random behavior in the system and time-variation in the input signal and noise spectra.

Under the assumptions listed above, the mean square error

spectrum takes the form of

$$\overline{e^2}(\omega) = S^2 + (S^2+N^2)M^2 - 2S^2 M \cos(\alpha). \quad (18)$$

The expected mean square error spectrum is therefore

$$E[\overline{e^2}(\omega)] = S^2 + (S^2+N^2)E[M^2] - 2S^2 E[M\cos(\alpha)]. \quad (19)$$

The closed loop gain (M) and phase shift (α) are assumed to be functions of some system parameter, τ, which exhibits some random behavior. The two Expected Values are generally evaluated by integration, for example:

$$E[M^2(\tau)] = \int_{\text{all } \tau} M^2(\tau)\cdot p(\tau)\cdot d\tau,$$

where $p(\tau)$ is the probability density function of τ. If the system environment causes τ to take on any of several discrete values rather than vary continuously, or if continuous variation can be reasonably approximated by discrete values and a discrete probability density function, the integration is replaced by summation over all the random values of τ.

The problem of designing the optimum system under these assumptions may be resolved into that of determining the best design value for the average of the random parameter τ. The general conclusion illustrated by this application of the method and by the examples to follow is that a better system (having less mean square error) will result from this use of probability theory than would result if the random behavior of the parameter were neglected, its average value being considered as a nonrandom parameter. Intuitively, the improvement can be seen to result from the fact that the expected

mean square error expression contains system function averages, such as $E[M^2]$ and $E[M\,Cos(\alpha)]$, rather than averages of the parameter alone.

As an example, the servosystem composed of an integrator in the open loop, system (1) of figure 2, is considered. The uncorrelated first order signal amplitude spectrum and white noise used previously (figure 1) are assumed for this example also. The mean square error spectrum was formulated in equation (18) above. The Expected Value operation may be applied to the spectrum, as in equation (19), or to the integral of the spectrum over all real frequencies. The operations of integration and Expected Value may be interchanged (5b) under the conditions that the spectrum is always less than some constant and is integrable in the Riemann sense, and that the Expected Value of the spectrum is also integrable in the Riemann sense. It is generally more convenient to perform the integration of the spectrum over all real frequencies, by the method of residues, first. Otherwise the integration associated with the Expected Value operation may result in expressions which are much more difficult to integrate by residues or any other means. The mean square error is as given in equation (17) above. Its Expected Value is

$$E[\bar{e}^2] = \frac{\pi}{2}\,\frac{\beta^2}{\delta}\left\{1 - E\left[\frac{1}{1+\delta\tau}\right] + E\left[\frac{1}{\tau}\right]\frac{1}{\Sigma+\delta}\right\}. \qquad (20)$$

The time constant τ is assumed to exhibit a discrete random

variation such that it may be either of two values, each with a probability of 1/2. The two values may be written as $\tau_o + \zeta\tau_o$ and $\tau_o - \zeta\tau_o$ where τ_o is the average value of the randomly varying time constant, and 2ζ is the relative spread, which will be considered fixed. The design problem is the determination of the value of τ_o which minimizes the mean square error. The Expected Values are readily evaluated, for example:

$$E\left[\frac{1}{\tau}\right] = \frac{1}{2}\,\frac{1}{\tau_o+\zeta\tau_o} + \frac{1}{2}\,\frac{1}{\tau_o-\zeta\tau_o}\,.$$

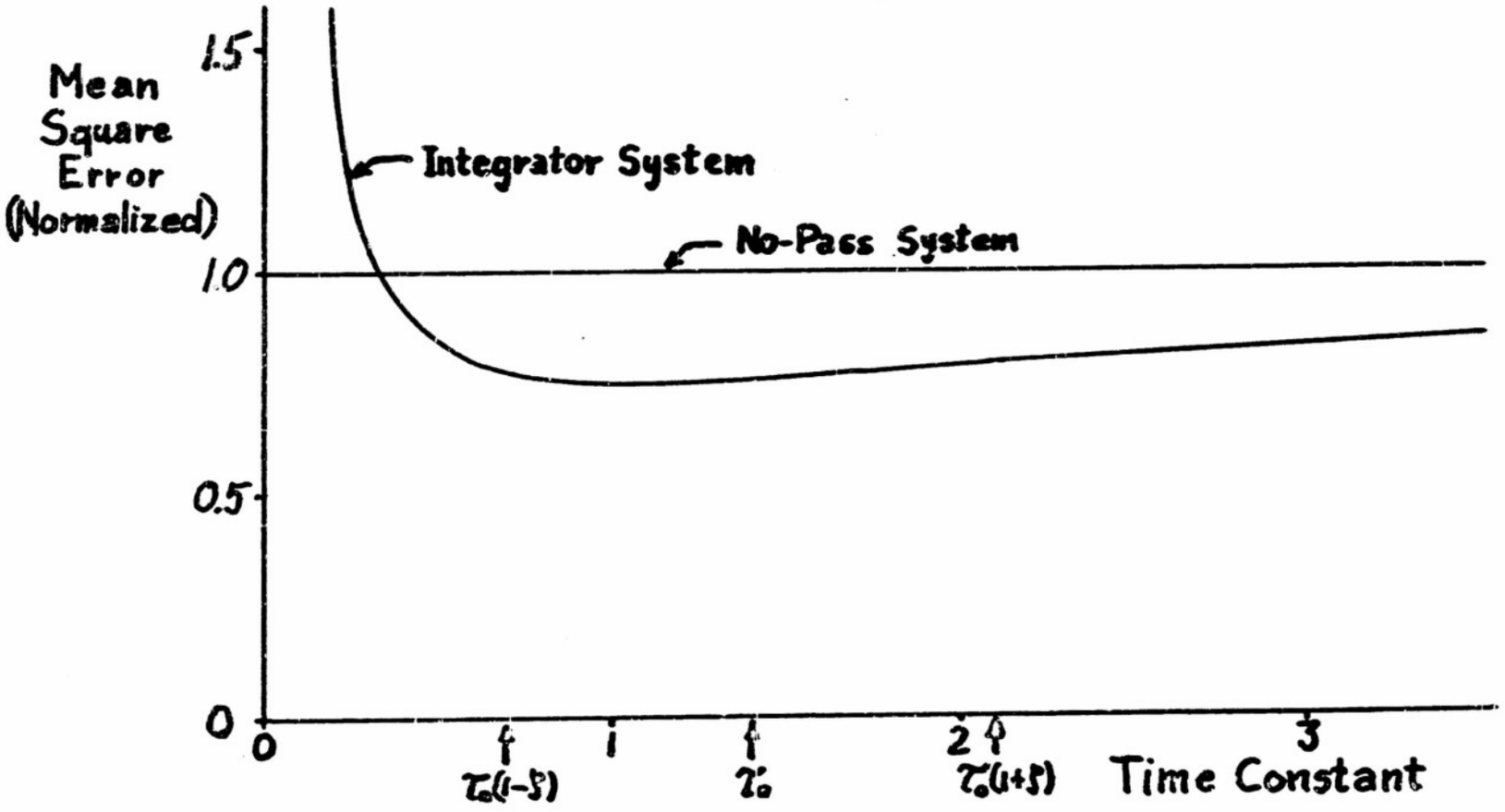
The expected mean square error may be conveniently written as

$$E\left[\overline{e^2}(\tau)\right] = \frac{1}{2}\,\overline{e^2}(\tau_o+\zeta\tau_o) + \frac{1}{2}\,\overline{e^2}(\tau_o-\zeta\tau_o). \qquad (21)$$

The minimization of this expression with respect to the average value, τ_o, of the time constant is accomplished by differentiation, subject to system stability considerations. However, a graphical presentation better illustrates the minimization process. A curve of the mean square error vs τ with $\Sigma = 2$ (equation 17) is presented in figure 4. As equation (21) indicates, the sum of the ordinates of that curve at $\tau = \tau_o + \zeta\tau_o$ and $\tau = \tau_o - \zeta\tau_o$ is to be minimized. Owing to the dissymmetry of the curve about its minimum, the optimum value of τ_o will be shifted from the minimum to some larger τ, the amount of the shift depending upon the value of ζ and the shape of the curve. For a particular example with $\Sigma = 2$ and $\zeta = 0.5$, the optimum value of τ_o is found to have shifted from the minimum of the curve at 1.0 to about 1.4.

Figure 4

Random System Parameter
Mean Square Error vs Time Constant
For $\Sigma = 2$



The most meaningful presentation of the improvement in system performance as a result of this method is to compare three values of the mean square error for the integrator system. For this example, these three values, normalized, are: a) 1.000 for an invariant integrator system with optimum $\tau = 1$, b) 1.067 for this system having the random time constant here assumed, but with $\tau_o = 1$ as a result of neglecting that random behavior in system design, and c) 1.041 for this system having the random time constant here assumed, and using the method herein presented to determine the optimum, $\tau_o = 1.4$. Thus, although the increase in system error due to randomness is small (6.7%), the application of this method lowered that increase by (67-41)/67 or 39%.

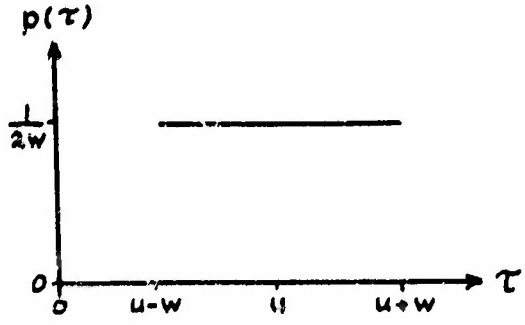
In the immediately previous example, a discrete probability density function was assumed to describe the random time-variations of the system time constant. In order that the use of continuous probability density functions may not be neglected, an example using such a density function is here presented. For simplicity of solution, this example will use the same system function and input spectra as were used for the example just preceding. The uniform probability distribution (4b) is chosen in this example to describe the time-varying, random behavior of the system time constant

It is defined as

$p(\tau) = 0$ for $\tau < u-w$,

$p(\tau) = 1/2w$ for $u-w \le \tau \le u+w$,

$p(\tau) = 0$ for $\tau > u+w$.



It is sketched at the right, where $\underline{u}$ is seen to be the average value of τ, and $2w$ is the extreme spread of the random variations of τ. The variance of τ, $E[(\tau-u)^2]$, is $w^2/3$. The optimization problem is to determine the optimum value of $\underline{u}$ in terms of the input spectra and of the spread, $\underline{w}$. The expected mean square error for the preceding example as given in equation (20) applies here also. It is repeated below for reference

$$E[\overline{e^2}] = \frac{\pi}{2}\frac{\beta^2}{\delta}\left\{1 - E\left[\frac{1}{1+\delta\tau}\right] + \frac{1}{\delta\Sigma^2}E\left[\frac{1}{\tau}\right]\right\} \qquad (20)$$

The evaluation of the Expected Values is as follows:

$$E\left[\frac{1}{\tau}\right] = \int_{all} \frac{1}{\tau}\cdot p(\tau)\cdot d\tau = \frac{1}{2w}\int_{u-w}^{u+w}\frac{1}{\tau}\cdot d\tau = \frac{1}{2w}\ln\left(\frac{u+w}{u-w}\right),$$

$$E\left[\frac{1}{1+\delta\tau}\right] = \frac{1}{2w\delta}\ln\left(\frac{1+\delta(u+w)}{1+\delta(u-w)}\right).$$

Substitution of these values in equation (20) yields

$$E[\overline{e^2}] = \frac{\pi}{2}\frac{\beta^2}{\delta}\left\{1 - \frac{1}{2w\delta}\ln\left(\frac{1+\delta(u+w)}{1+\delta(u-w)}\right) + \frac{1}{2w\delta\Sigma^2}\ln\left(\frac{u+w}{u-w}\right)\right\}. \qquad (22)$$

Differentiation of this expression, to determine the optimum $\underline{u}$, yields

$$\frac{\partial E[\overline{e^2}]}{\partial u} = \frac{\pi}{2}\frac{\beta^2}{\delta^2}\left\{\frac{\delta^2(\Sigma^2-1)u^2 - 2\delta u - \delta^2w^2(\Sigma^2-1) - 1}{\Sigma^2[(1+\delta u)^2 - \delta^2w^2][u^2-w^2]}\right\}.$$

Equating this derivative to zero, noting that system stability requires that $w < u$ so that τ is never negative, results in the roots $u = \dfrac{1 \pm \sqrt{\delta^2w^2(\Sigma^2-1)^2 + \Sigma^2}}{\delta(\Sigma^2-1)}$,

For the case of no randomness in τ , (w = o), the roots are $u = \frac{1}{\delta(\Sigma-1)}$ and $\frac{-1}{\delta(\Sigma+1)}$ which check with those obtained for the invariant integrator system treated earlier. The optimum value of the average of the randomly varying time constant is $u = \frac{1+\sqrt{\delta^2 w^2(\Sigma^2-1)^2+\Sigma^2}}{\delta(\Sigma^2-1)}$ The other root specifies a negative time constant and hence an unstable system. The optimum value of u above is only of value for $\Sigma > 1$. For $\Sigma < 1$, the average time constant is always negative and sometimes complex. As discussed for the invariant case, when $\Sigma \leq 1$ no integrator system can be an improvement over a no-pass system.

As an illustration of the results, consider the case of $\Sigma = 2$ and w = 0.5. The optimum value of the time constant has shifted from the invariant system value of 1.0 to a value of 1.17 for this time-varying case. The best means of presentation of the improvement in system performance resulting from the use of this method is to compare three values of the mean square error of the integrator system. These values are, normalized, a) 1.000 for an invariant integrator system with optimum $\tau = 1$, b) 1.016 for the system having the random time constant as described, but with u = 1 as a result of neglecting that random behavior when optimizing the system, and c) 1.011 for the system having the random time constant as described, and using the method presented herein to determine the optimum u = 1.17. Thus, although the increase in system error due to randomness is small (1.1%),

the application of this method reduced that increase by (16-11)/16 or 31%.

Other evaluations of the two examples presented, as well as different cases of this general type, indicate similar results. The increase in system mean square error due to such random behavior is not great, but this use of the Expected Value operation to determine the true optimum value for the average of that randomness recovers typically of the order of magnitude of a third of the increase.

The use of probability concepts was shown above to be necessary for the optimum design of those systems which are in essence similar to the examples treated. It is of interest to note that the results of those examples are of more general application than indicated by the particular problem stated. As an illustration of the generality of application, the results are here interpreted to be the solutions of two different problems. The first is to indicate the greatest range of random variations that can be tolerated in the value of a parameter without increasing the system mean square error appreciably. A possible application is the specification of the largest thermal coefficient of a parameter that could be tolerated. The second problem is to optimize the average of the performances of a large number of systems which are the same except that their construction involves one or more elements having values within some manufacturing tolerance. The proper nominal (average) value of the element and the tolerance required are the desired

results. That the method presented herein can determine such results has been adequately shown.

The cases described and carried to completion in this part of the dissertation have covered only a very few of the possible sets of conditions on random time-variations in system parameters and input spectra. The method presented herein is readily adaptable to much more complex problems with dependent as well as independent random variations.

CONCLUSIONS

This dissertation has presented an analytical method for the optimum design of servosystems characterized by random, time-varying parameters. The inputs to the systems are described by signal and noise voltage spectra which also are permitted to be time-varying. The method of optimization is limited in application to linear systems, and it uses the criterion of minimum mean square error. A spectrum of mean square error is derived and is shown to be of a form which is different from, although equivalent to, that commonly found in the literature. The form derived herein shows promise of easier application to some practical problems. Two procedures are introduced for the application of the mean square spectrum to the optimization problem. Each is illustrated by application to several examples. A formulation of the general solution for the more realistic, second procedure is given.

The method presented is not restricted in application to servo or feedback systems. It may be as readily applied to any linear system or network, especially those containing random, time-varying parameters or input spectra. The most important use of the method is felt to lie in its application to problems that are characterized by randomness in the parameters of the system or network involved.

FUTURE INVESTIGATIONS

The following suggestions indicate a few of the opportunities for future work in this field. An analytical investigation of the limitation on the rate of random time-variation, as described qualitatively above, would be a desirable extension of the work contained herein. That investigation could lead to a further extension of the method to cover random time-variations more rapid than the limitation now permits. Another extension of this work would comprise the solution of the general formulation given above for some more general, physically realizable system transfer functions.

A closely allied problem is the classic one of the determination of the size of a statistical sample of a random signal which is required to establish its spectral density or autocorrelation function to within some specified accuracy.

BIBLIOGRAPHY

1. Theory of Servomechanisms, M.I.T. Radiation Laboratory Series, Volume 25, H.M. James, N.B. Nichols, and R.S. Phillips, McGraw-Hill Book Company, Inc., 1947.

   a. Section 6.7
   b. Chapters 7 and 8
   c. Section 7.2

2. Servomechanisms and Regulating System Design, Volume I, H. Chestnut and R.W. Mayer, John Wiley and Sons, Inc., 1951.

   a. pp. 297-302
   b. p. 61

3. "Separating Information from Noise", Otto J.M. Smith, Transactions of the I.R.E. Professional Group on Circuit Theory, December 1952, p. 83

4. Introduction to the Theory of Statistics, A.McF. Mood, McGraw-Hill Book Company, Inc., 1950

   a. Chapter 5
   b. Chapter 6

5. Foundations of the Theory of Probability, A.N. Kolmogorov, Chelsea Publishing Co., 1950

   a. Chapter 4, Section 2, p. 39
   b. Chapter 4, Section 5, p. 44

6. Extrapolation, Interpolation, and Smoothing of Stationary Time Series, Norbert Wiener, Cambridge Technology Press of M.I.T., 1949, p. 95.

APPENDIX A

EXPERIMENTAL CHECK

This appendix contains a description of an experimental check on the use of the Expected Value operation in the method of this dissertation. The solutions of two examples of random time-variation in a system parameter are obtained analytically and compared with experimental results. The system chosen for these two examples is a simple lag circuit. Its time constant is made to vary discretely as was assumed in certain of the cases treated in the body of the dissertation. Slow, periodic switching from one value of the time constant to another is the special random time-variation chosen, for ease of experimental simulation. The input to this system is a single sine wave for the first example and a random noise spectral density for the second example.

These two examples constitute a sufficient check on the Expected Value operation. Statistically described signals contaminated with noise are not used because they are not essential to this check. In the paragraphs to follow, the general solutions of the dissertation are adapted to these examples and their solutions are presented. The experiments are described, and their results are compared with the analytical solutions.

Single Sine Wave Input

Analytical Solution. The system function for the simple lag circuit may be written as

$$Me^{j\alpha} = \frac{1}{1+j\omega\tau} = \frac{1}{\sqrt{1+\omega^2\tau^2}}\, e^{-j\tan^{-1}(\omega\tau)}$$

The input is $Sin(\omega t)$.

The steady state output is then

$$e = \frac{1}{\sqrt{1+\omega^2\tau^2}}\, Sin(\omega t - \tan^{-1}(\omega\tau)).$$

The mean square output is therefore

$$\overline{e^2} = \frac{\omega}{2\pi}\int_0^{2\pi/\omega} e^2\, dt = \frac{1}{2}\frac{1}{1+\omega^2\tau^2}. \qquad (23)$$

The system time constant, τ, is assumed to take on two values, τ_1, and τ_2, with equal probabilities. Thus the probability density function, $p(\tau)$, has the values $p(\tau_1) = 1/2$ and $p(\tau_2) = 1/2$. The expected mean square output is then

$$E[\overline{e^2}] = \sum_{i=1,2} \overline{e^2}(\tau_i)\cdot p(\tau_i),$$

$$E[\overline{e^2}] = \frac{1}{2}\left\{\frac{1}{2}\frac{1}{1+\omega^2\tau_1^2} + \frac{1}{2}\frac{1}{1+\omega^2\tau_2^2}\right\}.$$

This solution assumes that the frequency of switching between τ_1 and τ_2 is slow enough that any transient effects are negligible. Using τ_1 as a base and setting $\tau_2 = k\tau_1$, the expected mean square output is

$$\mathcal{E}(k) \triangleq E[\overline{e^2}] = \frac{1 + \frac{1+\omega^2\tau_1^2}{1+\omega^2\tau_1^2k^2}}{4(1+\omega^2\tau_1^2)}. \qquad (24)$$

The solution above may be conveniently presented as a

ratio of the above result to that for no random variation, i.e. for $K=1$ and therefore $\tau_2 = \tau_1$.

$$\frac{\mathcal{E}(K)}{\mathcal{E}(1)} = \frac{1 + \frac{1+\omega^2\tau_1^2}{1+\omega^2\tau^2K^2}}{2} \qquad (25)$$

This ratio is plotted in figure 5, which follows this appendix, for a particular value of $\omega\tau_1$ and as a function of K .

Experimental Check. The experimental setup for this first example consists of the equipment described below and is illustrated in figure 6, which follows this appendix. A signal generator supplies the single sine wave input to the system which is an R-C lag circuit. The system is isolated before and after by cathode follower circuits, which are formed on the Operational Manifold (George A. Philbrick Researches, Inc.). The capacitance of the circuit is varied discretely by adding and removing additional capacitance, using a motor driven switch. The motor, the gear train drive, and the cam-operated micro-switch are conveniently assembled using the Belock Instrument Corporation Servo Components available. The output of the system, fed through a cathode follower, is coupled to the heater coil of a 22J Western Electric thermocouple. The resulting squared and averaged output voltage is read on a light bear type galvanometer.

The system was first checked as an invariant system against the known attenuation versus frequency characteristic

of a simple lag circuit and was found to be in good agreement. The experimental procedure consisted essentially of selecting a suitable base value for $\omega\tau$, (this was chosen for 100cps, R = 10K, and C = 0.4 mfd.), varying the value of the other capacitor, and recording the galvanometer deflections for three situations. These are each of the two positions of the switch and the slowly switching condition. The experimental ratios corresponding to equation (25) are plotted on the analytical curve of that ratio, figure 5, as circles. They are seen to lie within 2. % of the analytical solution.

Spectral Density Input

Analytical Solution. The system function is again that of the simple lag circuit.

$$Me^{j\alpha} = \frac{1}{1+j\omega\tau}$$

The input spectral density chosen is that expected from the General Radio Random Noise Generator, Type 1390-A. The amplitude spectrum may be written as

$$N(\omega) = \frac{\eta\omega}{\sqrt{\omega^2+\rho^2}}$$

where η is the rms amplitude at high frequency and ρ is the lower out-off (-3db) frequency. The mean square output spectrum, adapted from equation (7), is

$$\overline{e^2}(\omega) = N^2(\omega)\cdot M^2(\omega) = \frac{\eta^2\omega^2}{\omega^2+\rho^2}\cdot\frac{1}{1+\omega^2\tau^2}.$$

The mean square output is the integral of that spectrum.

$$\overline{e^2} = \int_0^\infty \overline{e^2}(\omega) \cdot d\omega = \frac{\pi}{2} \frac{\eta^2}{\tau(1+\rho\tau)} \qquad (26)$$

The expected mean square output, obtained in the same manner as in the preceding example, is thus

$$E[\overline{e^2}] = \frac{\pi}{2}\eta^2 E\left[\frac{1}{\tau(1+\rho\tau)}\right] = \frac{\pi}{2}\eta^2\left\{\frac{1}{2}\frac{1}{\tau_1(1+\rho\tau_1)} + \frac{1}{2}\frac{1}{\tau_2(1+\rho\tau_2)}\right\}.$$

Setting $\tau_2 = k\tau_1$ gives the simpler form of the result

$$\mathcal{E}(k) \triangleq E[\overline{e^2}] = \frac{\pi}{2}\eta^2 \frac{1}{2} \frac{1 + \frac{\tau_1(1+\rho\tau)}{k\tau_1(1+\rho\tau_1 k)}}{\tau_1(1+\rho\tau_1)}$$

A ratio is formed as in the single sine wave case above.

$$\frac{\mathcal{E}(k)}{\mathcal{E}(1)} = \frac{1 + \frac{(1+\rho\tau_1)}{k(1+\rho\tau_1 k)}}{2} \qquad (27)$$

This ratio is shown as a function of k in figure 7, for a particular value of τ_1 and of ρ .

Experimental Check. The experimental setup for this example differs from that of the preceding example in that the Random Noise Generator is substituted for the sine wave generator, and an amplifier and another cathode follower (both formed on the Operational Manifold) are inserted before the thermocouple heater to compensate for the low output level available from the Random Noise Generator. This setup is also shown in figure 6.

The procedure followed here is the same as that of the preceding example. The value of ρ is determined for the analytical curve by using the galvanometer deflections for two values of τ with the system invariant. The ratio of

the deflections and the two values of τ are substituted in the analytical expression for the ratio. It is

$$\frac{\overline{e^2}(\tau_1)}{\overline{e^2}(\tau_2)} = \frac{\tau_2(1+\rho\tau_2)}{\tau_1(1+\rho\tau_1)} \triangleq R.$$

Therefore,

$$\rho = \frac{\tau_2 - R\tau_1}{R\tau_1^2 - \tau_2^2}$$

The experimental ratios which correspond to equation (27) are calculated from the galvanometer deflections for the base time constant (τ_1) alone and for the slowly switching condition of the system time constant. These experimental ratios are plotted as circles on the analytical curve of that ratio, figure 7. The experimental values are seen to lie within 1.5% of the analytical curve. The Random Noise Generator does not quite have a true first order characteristic at low frequencies. A first order filter is coupled on the output of the generator to correct this condition.

These two experiments indicate a good agreement between the experimental results and the analytical solutions using the Expected Value operation. It is concluded that the use of this probability theory operation is certainly valid, with one limitation. The rate of the random variation must be slow enough that any resulting transients will have negligible effects on the mean square value of the output.

Figure 5

Comparison of Experimental Points and Analytical Curve

Mean Square Error Ratio : $\mathcal{E}(k)/\mathcal{E}(1)$
versus the Random Variation Factor : k

Single Sine Wave Input

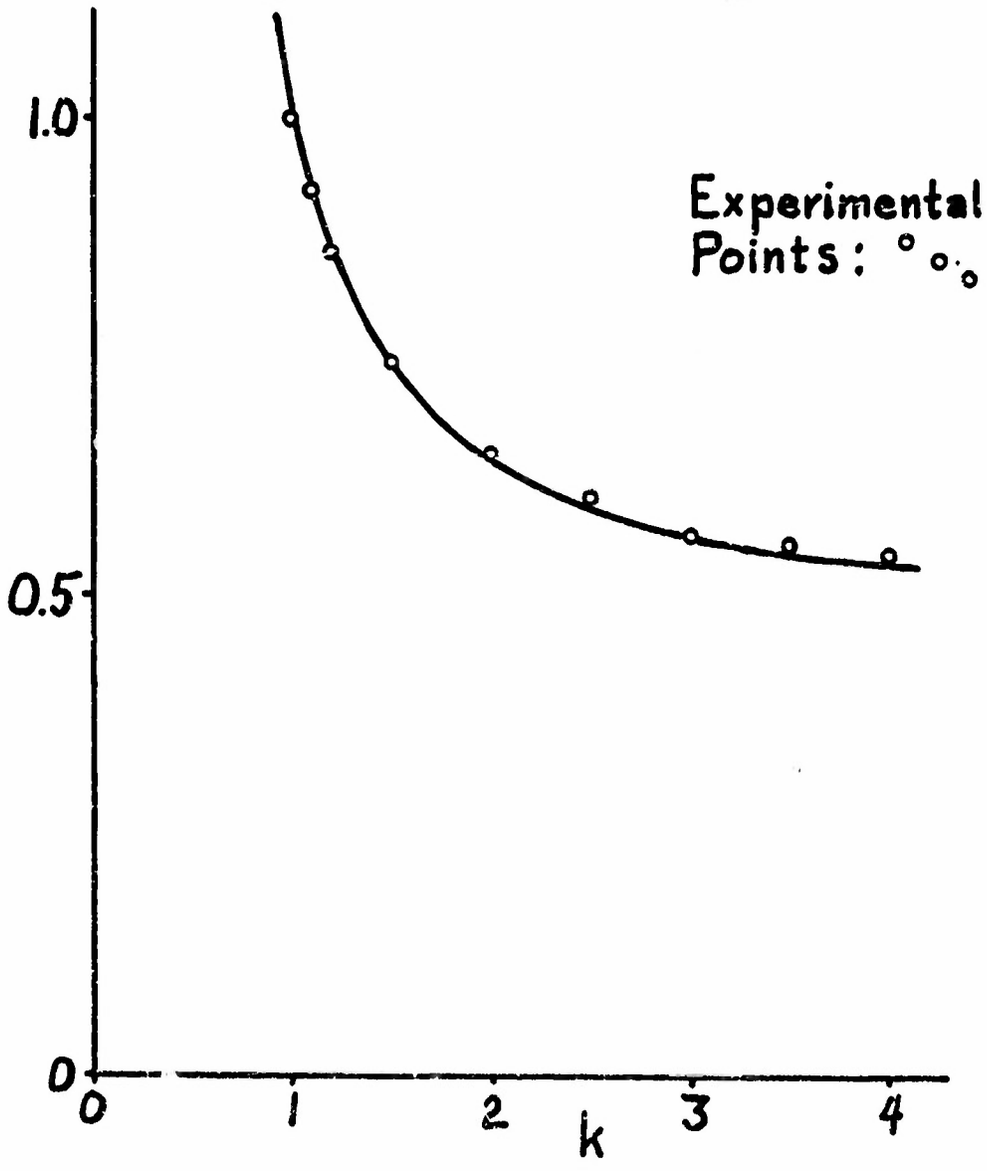
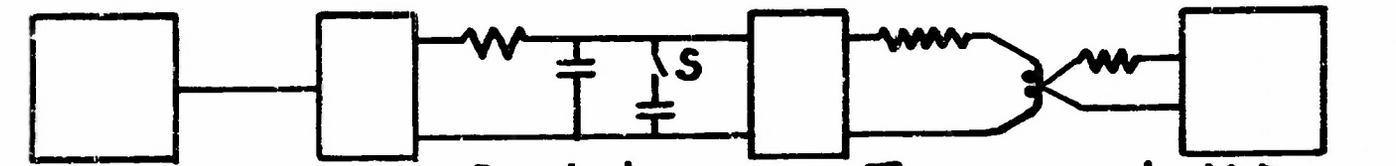


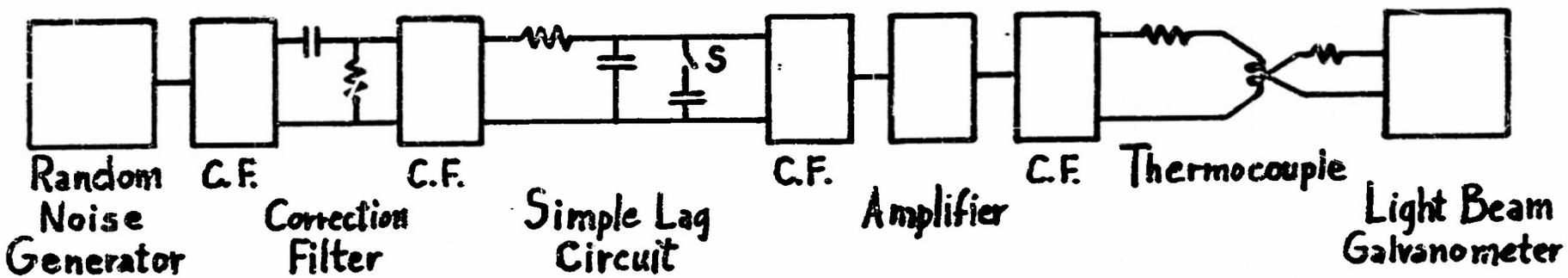
Figure 6

Experimental Setups for a Check on the Expected Value Operation

Single Sine Wave Input



Spectral Density Input



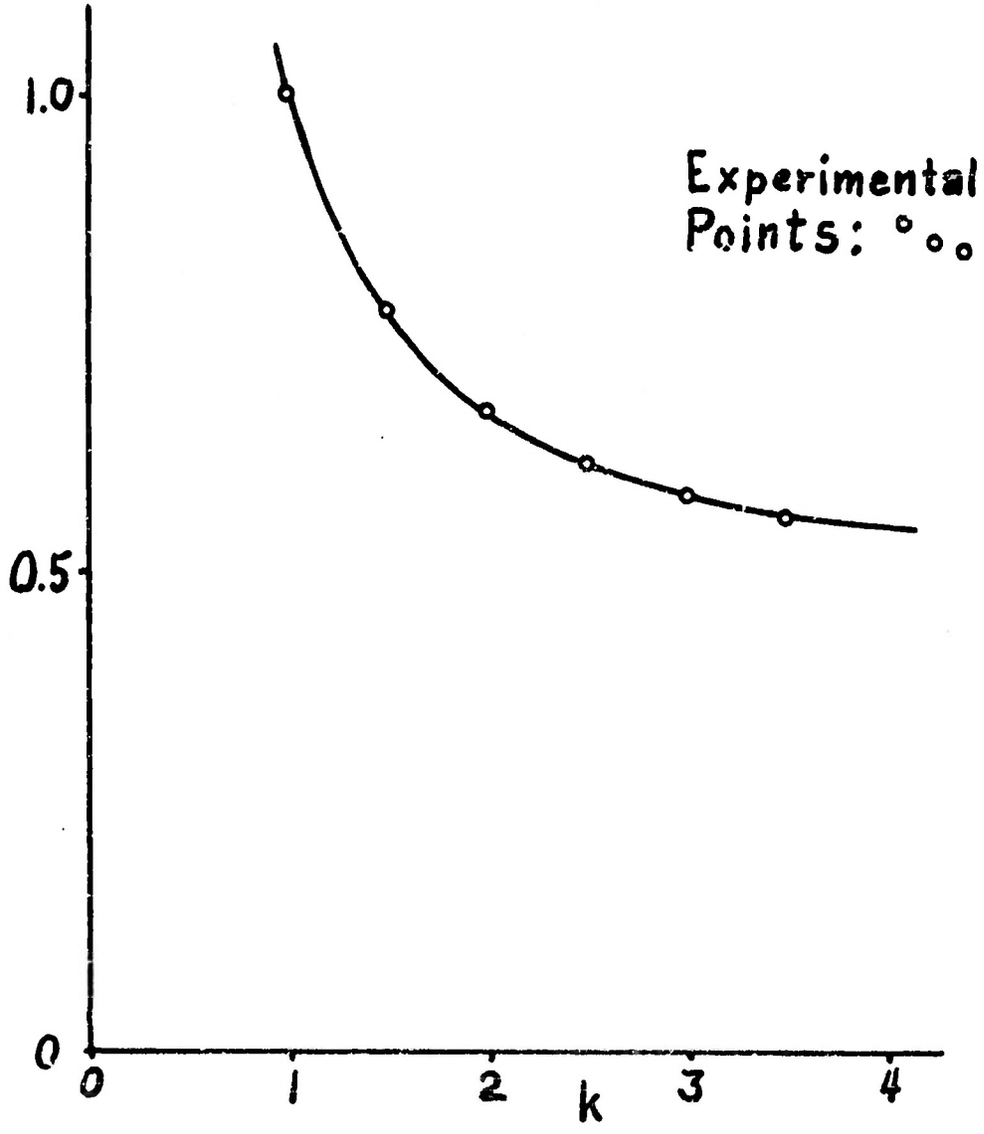
Key: C.F. is Cathode Follower, S is the motor-driven micro-switch.

Figure 7

Comparison of Experimental Points and Analytical Curve

Mean Square Error Ratio : $\mathcal{E}(k)/\mathcal{E}(1)$ versus the Random Variation Factor k

Spectral Density Input



APPENDIX B

OPTIMIZATION OF A FIRST ORDER SYSTEM

This appendix describes the optimization procedure for system (3) as shown in figure 2. The results of this optimization were discussed in the body of this dissertation and were presented in figure 3. The open loop of this system consisted of a simple lag plus an ideal amplifier. The closed loop system transfer function is

$$M e^{j\alpha} = \frac{K}{K+1+j\omega\tau} ,$$

$$M = \frac{K}{\sqrt{(K+1)^2 + \omega^2\tau^2}} , \quad \alpha = \cos^{-1}\frac{K+1}{\sqrt{(K+1)^2+\omega^2\tau^2}} .$$

The signal and noise are assumed to be uncorrelated. They were shown in figure 1 and are

$$S(\omega) = \beta / \sqrt{\omega^2+\delta^2} , \quad N(\omega) = \eta .$$

The signal to noise ratio is as defined before: $\Sigma = \beta/\delta\eta$.

The proper form of the mean square error spectrum is equation (16), repeated here for reference.

$$\overline{e^2}(\omega) = S^2 + (S^2+N^2)M^2 - 2S^2 M \cos(\alpha). \tag{16}$$

Substitution of the assumed system function and input spectra yields

$$\overline{e^2}(\omega) = \frac{\beta^2}{\omega^2+\delta^2} + \left(\frac{\beta^2}{\omega^2+\delta^2} + \eta^2\right)\frac{K^2}{(K+1)^2+\omega^2\tau^2} - 2\frac{\beta^2}{\omega^2+\delta^2}\frac{K(K+1)}{(K+1)^2+\omega^2\tau^2} .$$

The mean square error is the integral of this spectrum over all real, positive frequencies. By the method of residues,

$$\overline{e^2} = \frac{\pi}{2}\left\{\frac{\beta^2}{\delta}\frac{1+(K+1)\delta\tau}{(K+1)(K+1+\delta\tau)} + \frac{\eta^2K^2}{\tau(K+1)}\right\}.$$

It may be written in terms of the signal to noise ratio, Σ, and a dimensionless time constant, $T=\delta\tau$, as

$$\overline{e^2} = \frac{\pi}{2}\frac{\beta^2}{\delta\Sigma^2}\left\{\Sigma^2\frac{1+(K+1)T}{(K+1)(K+1+T)} + \frac{K^2}{T(K+1)}\right\}. \qquad (28)$$

It is this mean square error that is to be minimized by proper choice of the system time constant, τ, and the gain, K. Minimization requires that the partial derivatives of $\overline{e^2}$ with respect to T and K be equal to zero. Setting the first partial equal to zero yields

$$\frac{\partial\overline{e^2}}{\partial T} = \frac{\pi}{2}\frac{\beta^2}{\delta\Sigma^2(K+1)}\left\{\Sigma^2\frac{K(K+2)}{(K+1+T)^2} - \frac{K^2}{T^2}\right\} = 0,$$

$$T = (K+1)K\frac{K \pm \Sigma\sqrt{K(K+2)}}{[\Sigma\sqrt{K(K+2)}-K][\Sigma\sqrt{K(K+2)}+K]}.$$

The value of T using the negative sign above yields

$$T_- = K(K+1)\frac{-1}{\Sigma\sqrt{K(K+2)}+K}.$$

which is always negative (for $K>0$) and there results an unstable system. The root using the positive sign is

$$T_+ = K(K+1)\frac{1}{\Sigma\sqrt{K(K+2)}-K}. \qquad (29)$$

Inspection of this root by substitution in the mean square error, equation (28), and by other means establishes it as the optimum time constant. As a check, this root reduces

to the optimum time constant for the simple lag system (system (2) of figure 2) when K is set equal to unity. The optimum time constant, equation (29) is substituted into the mean square error, equation (28), and the partial derivative with respect to K is equated to zero. There results a cubic in K ,

$$(1-\Sigma^2)K^3+(2-4\Sigma^2)K^2+(\Sigma^4-5\Sigma^2)K+2\Sigma^4=0,$$

which may be factored,

$$[(1-\Sigma^2)K-2\Sigma^2]\cdot[K^2+2K-\Sigma^2]=0,$$

to determine the three roots. These are

$$K_1=-1-\sqrt{\Sigma^2+1}\ ,\quad K_2=-1+\sqrt{\Sigma^2+1},\quad K_3=2\Sigma^2/(1-\Sigma^2).$$

In general, these roots indicate the maximum, minimum, and turning points of the function. Rather than form the second partial of $\overline{e^2}$ with respect to K to check these roots for a minimum of $\overline{e^2}$, they are substituted in $\overline{e^2}$ and the results are inspected. Substitution of the first root yields a negative value of $\overline{e^2}$, an impossible physical situation resulting from K being sufficiently negative and hence the system being unstable. The third root yields a mean square error which is always the same as that for a no-pass system. Substitution of the second root results in a value of $\overline{e^2}$ which is always less than that for a no-pass system. It is this second root that is the optimum gain within the limitations of system stability. The optimum time constant is thus equation (29) evaluated for K_2:

$$T=\frac{\sqrt{\Sigma^2+1}\ \left(-1+\sqrt{\Sigma^2+1}\right)}{\Sigma\sqrt{\left(\sqrt{\Sigma^2+1}+1\right)\left(\sqrt{\Sigma^2+1}-1\right)}+1-\sqrt{\Sigma^2+1}}=1.$$

The optimum system has a minimized mean square error of

$$\overline{e^2} = \frac{\pi}{2} \frac{\beta^2}{\delta \Sigma^2} 2 \left(-1 + \sqrt{\Sigma^2+1}\right) \tag{30}$$

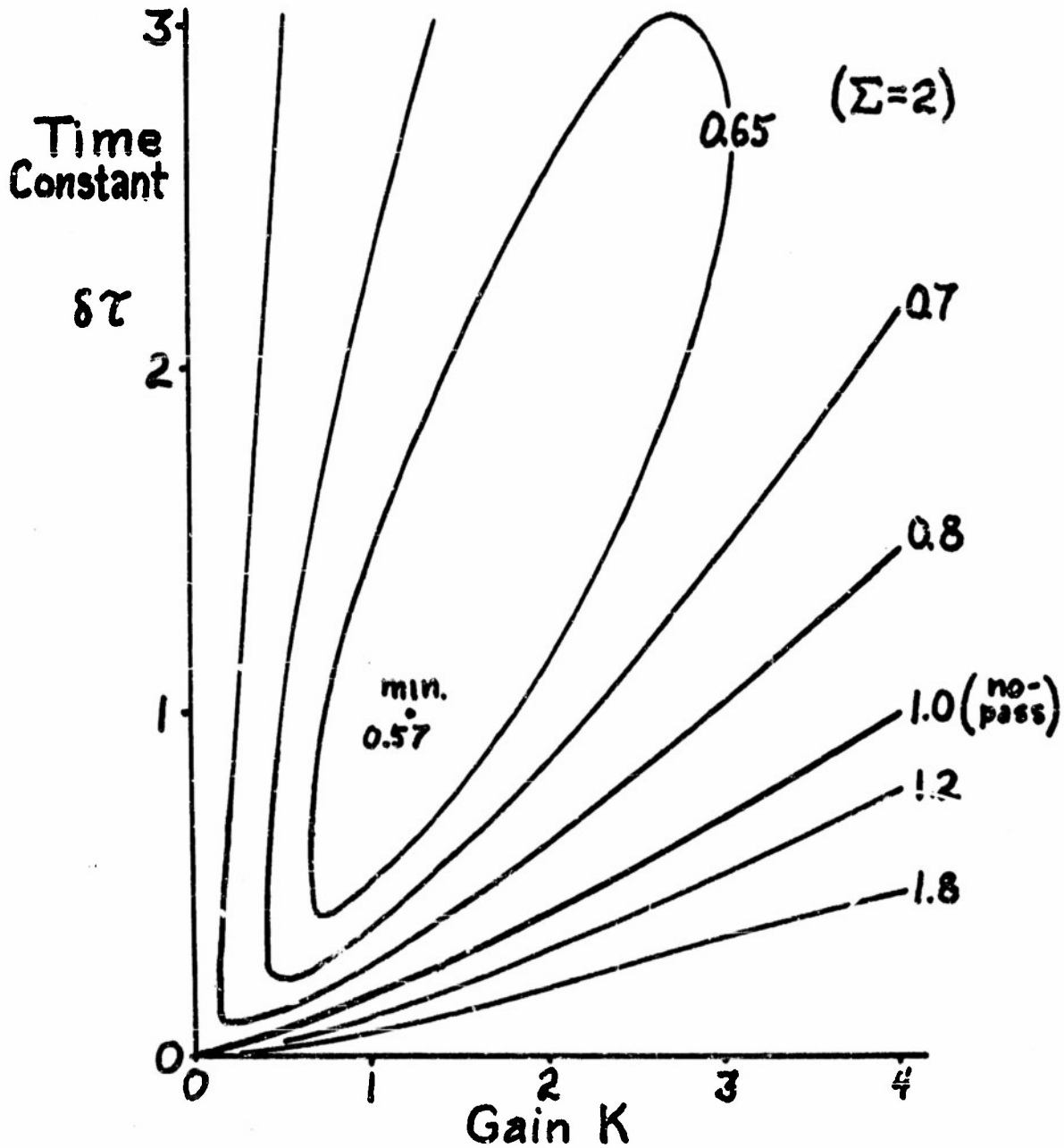
which results from substituting the optimum gain, $K = -1+\sqrt{\Sigma^2+1}$, and the optimum time constant, $\tau = 1/\delta$, in equation (28). It is this mean square error, equation (30), which is plotted in figure 3 above and is there discussed.

The surface generated by $\overline{e^2}$ as a function of K and T is of great value in visualizing this optimization process. For that reason, a contour map of that surface (for $\Sigma = 2$) is included here as figure 8.

Figure 8

Contour Map of the Mean Square Error Surface for System 3

Normalized to the mean square error of a no-pass system



APPENDIX C

OPTIMIZATION OF A SECOND ORDER SYSTEM

This appendix describes the optimization of the second order system described as the fourth invariant system in the body of the dissertation. It has a closed loop transfer function which is the product of two simple lags with time constants τ_1 and τ_2. It is written as

$$M e^{j\alpha} = \frac{1}{(1+j\omega\tau_1)(1+j\omega\tau_2)} .$$

The signal and noise inputs are assumed to be uncorrelated and are described by their amplitude spectra as shown in figure 2 above and as written below.

$$S(\omega) = \beta / \sqrt{\omega^2+\delta^2} \quad , \quad N(\omega) = \eta .$$

The mean square error spectrum for these assumptions is equation (16), repeated here for convenience.

$$\overline{e^2}(\omega) = S^2 + (S^2+N^2)M^2 - 2S^2 M \cos(\alpha). \qquad (16)$$

Substitution of the system gain and phase shift and the signal and noise amplitude spectra yields

$$\overline{e^2}(\omega) = \frac{\beta^2}{\omega^2+\delta^2} + \left(\frac{\beta^2}{\omega^2+\delta^2}+\eta^2\right)\frac{1}{(1+\omega^2\tau_1^2)(1+\omega^2\tau_2^2)} - 2\frac{\beta^2}{\omega^2+\delta^2}\cdot\frac{1-\omega^2\tau_1\tau_2}{(1+\omega^2\tau_1^2)(1+\omega^2\tau_2^2)} .$$

The mean square error is formed by integrating this spectrum over all real, positive frequencies by the method of resi-

dues.

$$\overline{e^2} = \frac{\pi}{2}\left\{\frac{\beta^2}{\delta} + \frac{(\delta\tau_1\delta\tau_2 - \delta\tau_1 - \delta\tau_2)\beta^2 + (1+\delta\tau_1)(1+\delta\tau_2)\eta^2\delta^2}{\delta(\delta\tau_1+\delta\tau_2)(1+\delta\tau_1)(1+\delta\tau_2)}\right\}.$$

This may be written in terms of the signal to noise ratio, $\Sigma = \beta/\delta\eta$, and two dimensionless time constants, $T_1 = \delta\tau_1$, and $T_2 = \delta\tau_2$, as

$$\overline{e^2} = \frac{\pi}{2}\frac{\beta^2}{\delta\Sigma^2}\left\{\Sigma^2 + \frac{(T_1T_2 - T_1 - T_2)\Sigma^2 + (1+T_1)(1+T_2)}{(T_1+T_2)(1+T_1)(1+T_2)}\right\}. \qquad (31)$$

Equating to zero the partial of $\overline{e^2}$ with respect to T_1 yields

$$T_1 = \frac{1 - T_2(\Sigma^2-1) \pm \Sigma\sqrt{(2T_2-1)[T_2^2(\Sigma^2+1)-1]}}{(\Sigma^2-1) - T_2(\Sigma^2+1)} \qquad (32)$$

For the special case of $T_2 = 0$, these roots check with those derived for the integrator system (system (1) of figure 3). Owing to the complexity of the roots, a graphical-analytical approach is used. Specific values of Σ are substituted in equations (32). Those equations are plotted in the T_1 versus T_2 plane and give the curves along which the partial derivative of $\overline{e^2}$ with respect to T_1 is zero. The similar curves along which the partial of $\overline{e^2}$ with respect to T_2 is zero are also plotted. These latter curves are obtained by interchanging T_1 and T_2 in equation (32), making use of the symmetry of $\overline{e^2}$ with respect to T_1 and T_2. The intersections of these curves are expected to indicate the conditions for minimum and maximum $\overline{e^2}$, and possibly a saddle point of the surface of $\overline{e^2}$ as a function of T_1 and T_2. Only the intersections in the first quadrant are of interest as that is the region of

positive time constants and hence system stability. Evaluation of $\overline{e^2}$ at those intersections and at adjacent points reveals that the intersections in the first quadrant indicate maxima and saddle points of $\overline{e^2}$. The minimum values of $\overline{e^2}$ are to be found at intersections in the other quadrants, outside the region of system stability. Therefore the restricted minimum of $\overline{e^2}$ must lie on the boundaries of the first quadrant, i.e. $T_1 = 0$ or $T_2 = 0$. These boundaries correspond to the previously treated first order system having an integrator in the open loop. It was optimized in the body of the dissertation. The optimum second order system, for the assumptions of this case, is therefore a first order system. Optimization dictates that one time constant be zero and the other be as given by equation (32) with T_2 equal to zero and the positive sign used. That value is

$$T = 1/(\Sigma^2 - 1).$$

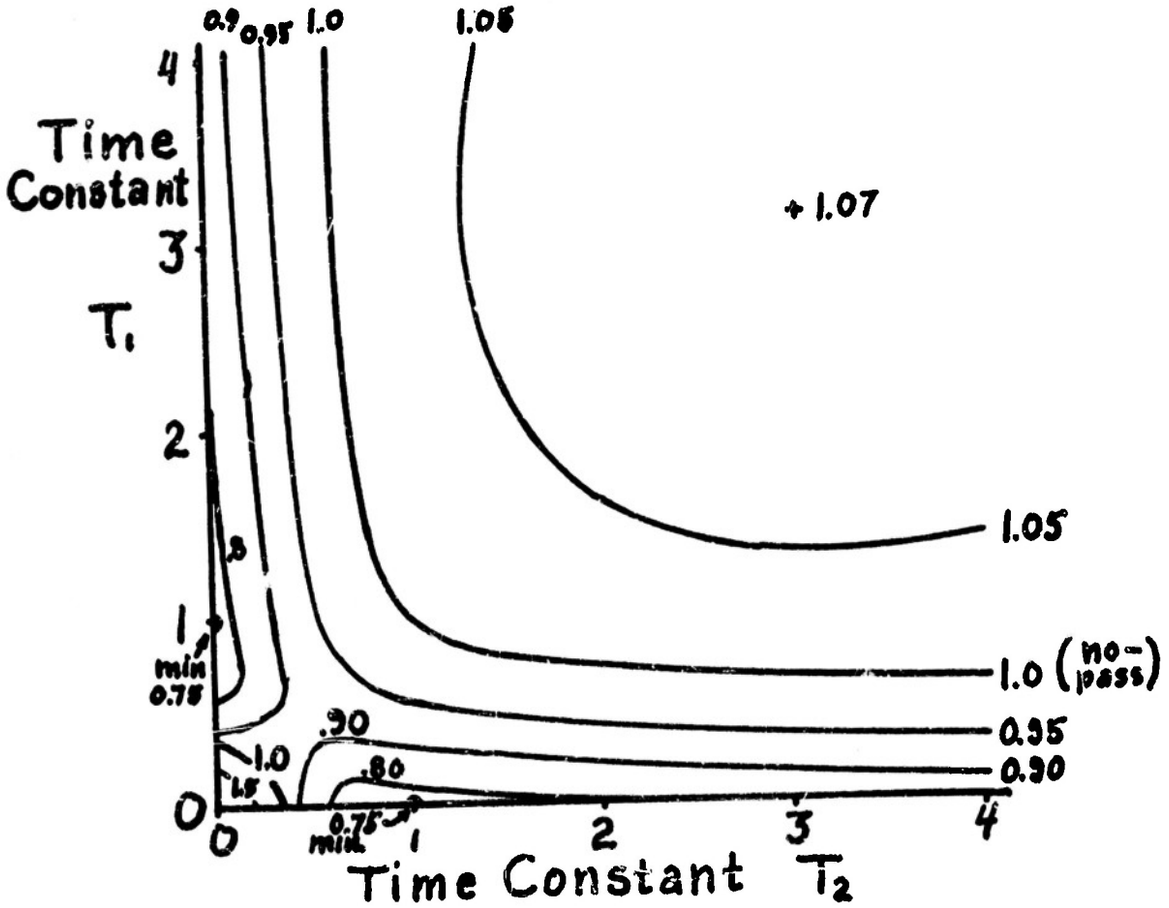
Consideration of the shape of the surface generated by $\overline{e^2}$ as a function of T_1 and T_2 is of great value in visualizing the situation. A contour map of that surface, for $\Sigma = 2$, is included here as figure 9. Investigation of the boundary between mean square error less than and greater than that for a no-pass system generally leads to a simpler polynomial than the general contour of the surface. It is also of value in visualizing the situation. Another use of the surface representation is that it presents the solutions of the restricted problem of one time constant being fixed

Figure 9

Contour Map of the Mean Square Error Surface for the Second Order System

Normalized to the mean square error of a no-pass system

$\Sigma=2$



in value. The optimum value of the remaining time constant is readily discovered by inspection of the proper section through that surface.